VOLUME 104 PART B NUMBER 16 JULY 1957

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 : INCORPORATED BY ROYAL CHARTER 1921

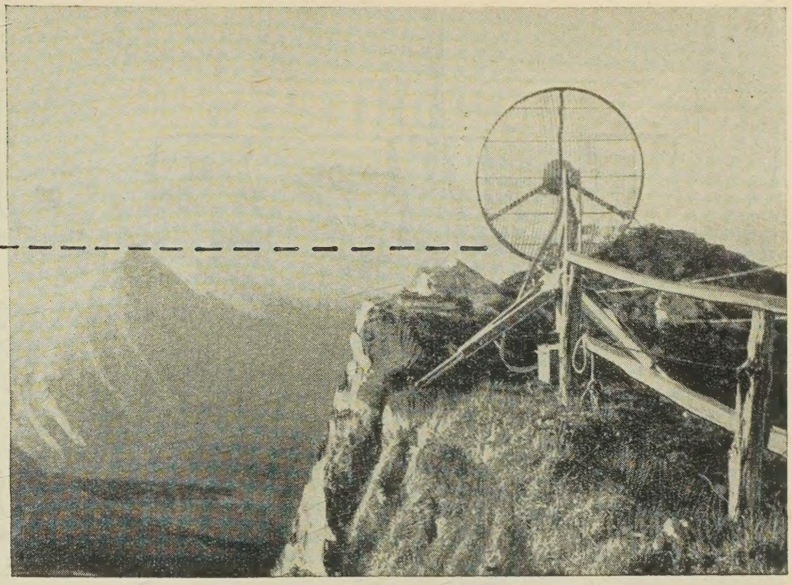
PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE · LONDON W.C.2

Price Ten Shillings

One of the E.M.I. Television Microwave Relay Links in operation in the Swiss Alps forms part of the Eurovision Network.



E·M·I

TELEVISION Relay Equipment

E.M.I. High Power Television Microwave Links are designed to provide a high quality video link between an outside broadcast unit and the main T.V. station or for projects such as the Eurovision Network. They are suitable for 405, 525 and 635 line systems, are portable and of robust construction. Range over unobstructed path at least 40 miles.

TYPE ML6B (3 Watt Klystron)

- Operating Frequency 4,400-4,800 Mc/s per sec.
- Available with both sound and vision channels.

TYPE ML4A (2 Watt Klystron)

- Operating Frequency 6,875-7,300 Mc/s per sec.
- Suitable for colour as well as monochrome transmissions.

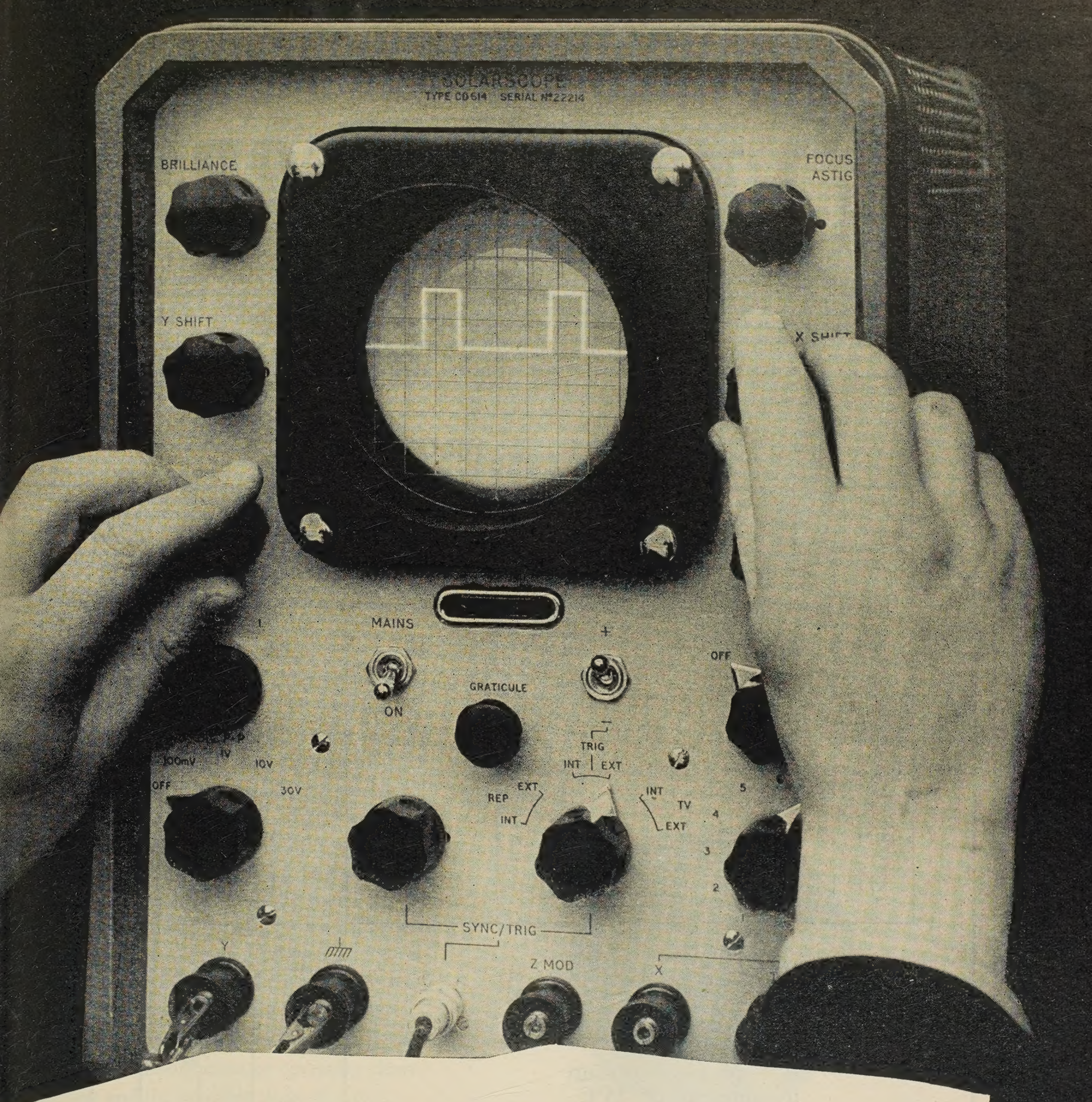
E.M.I... PIONEERS OF THE WORLD'S FIRST PUBLIC HIGH DEFINITION TELEVISION SERVICE IN 1936...E.M.I. PATENTED TELEVISION INVENTIONS ARE EXTENSIVELY USED UNDER LICENCE.

OTHER E.M.I. DEVELOPMENTS FOR TELEVISION TRANSMISSION
Studio Mixing and Programme Equipments, Camera and Film Channels, including Colour Film, Flying Spot and Photo-Conductive Scanners. Also Industrial Television.

For particulars of the above-mentioned equipments write to:—

E.M.I. ELECTRONICS LTD., VALVE DIVISION, HAYES, MIDDLESEX, ENGLAND
Tel: SOUthall 2468

EE92



High performance . . . wide applications . . . truly portable

In the Solarscope CD 614 we have included all the valuable features of heavier and more expensive oscilloscopes, while producing a truly portable instrument at an economical price. It is particularly suited for radio communication, radar, TV and applications involving pulse work and transient investigations.

BRIEF SPECIFICATION

NOMINAL BANDWIDTH
1 c/s — 9 Mc/s ± 1 Mc/s for 3 db down

SENSITIVITY CALIBRATION
By a 50 c.p.s. square wave

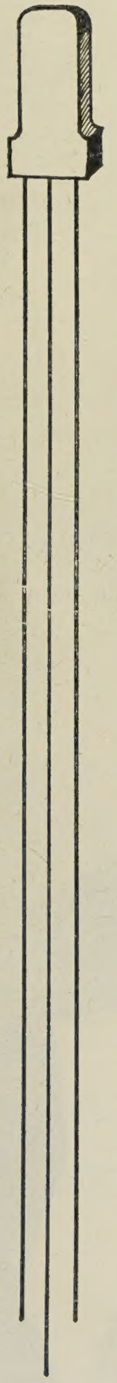
EXPANSION
10 diameters nominal

CALIBRATION
By 0·1 μS, 1μS, and 10 μS markers ± 5%

TIME BASE
10 c.p.s. — 200 Kc/s
Trigger from TV frame block

THE SOLARTRON ELECTRONIC GROUP LTD.

B T-H TRANSISTORS



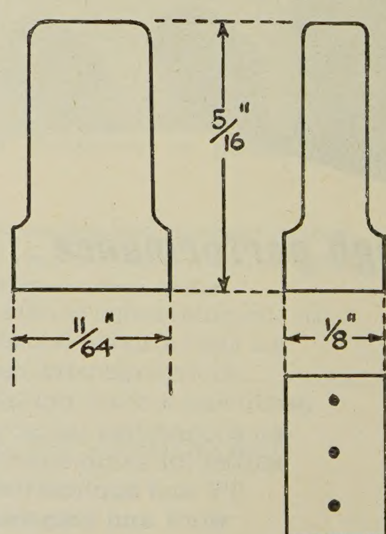
Pioneers in the field of crystal devices, British Thomson-Houston, by constant advances in manufacturing techniques, are able to supply transistors of unsurpassed quality at highly competitive prices.

AVAILABLE FROM STOCK

Type	Typical cut off frequency (Common Base) Mc/s	Current Amplification h'_{21} (Common Emitter)	Maximum Peak Collector Voltage Volts	Maximum Collector Current mA	Maximum Power Dissipation mW
GT 1	0·8	20	–15	–10	50
GT 2	0·9	40	–15	–10	50
GT 3	1·0	65	–15	–10	50
GT 11	4·0	30	–15	–10	25
GT 12	6·0	40	–15	–10	25
GT 13	9·0	50	–15	–10	25

Ratings given pertain to ambient of 25°C.

Maximum Storage Temperature 55°C.



BRITISH THOMSON-HOUSTON

THE BRITISH THOMSON-HOUSTON COMPANY LIMITED • RUGBY • ENGLAND

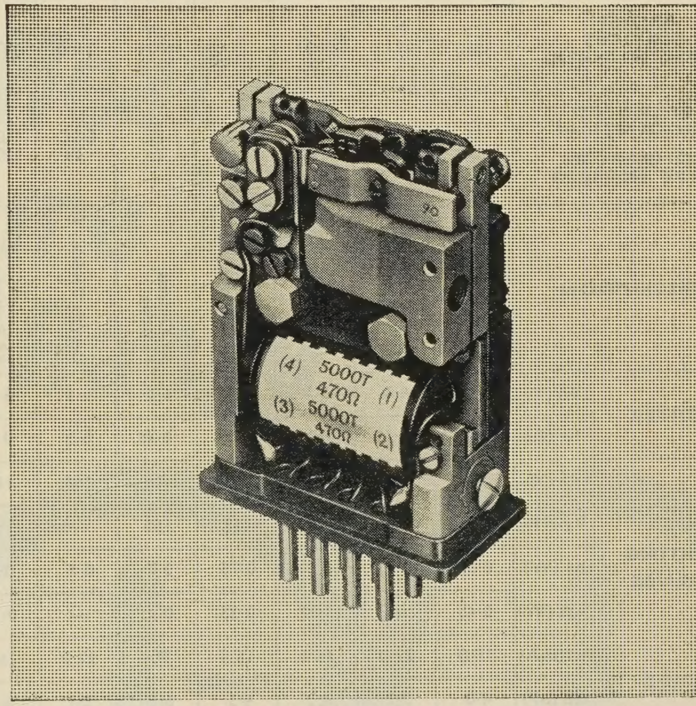
an A.E.I. Company

A5132

Now in production!

The Centre-Stable

TYPE 51M CARPENTER POLARIZED RELAY



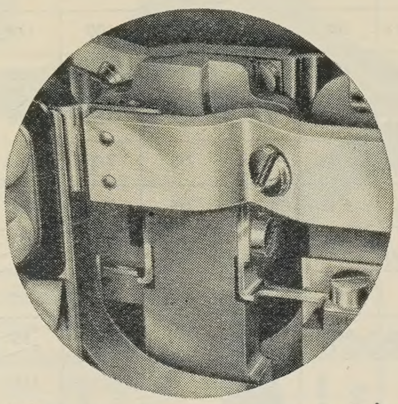
DIMENSIONS (excluding connecting tags and guide pins): HEIGHT: $2\frac{3}{8}$in. WIDTH: $1\frac{17}{32}$in. DEPTH: $2\frac{5}{32}$in.

This new highly sensitive *centre-stable** polarized relay is a development of the well-known Type 5 Carpenter Relay.

Like its companion—the *each-side-stable* Type 51A relay—the 51M's high performance is due largely to a new form of armature suspension, and the provision of variable slugs for adjusting the flux of the polarizing magnets thereby enabling the magnetic and mechanical parameters to be balanced with greater precision than has hitherto been possible.

With a sensitivity of approximately 10μW, the Type 51M relay is ideally suited for use as a sensing relay in Servo Systems, or as a detector in self balancing bridge circuits. Fitted with S.P.D.T. contact action, the relay is available in sealed or non-sealed forms.

For further details send for leaflet F.3526 which gives information of operating characteristics and the available range of coils.



Close-up of armature suspension —magnet removed

* *In the centre-stable form, the armature of the relay remains in a central position between the side-contacts when the relay is unenergized, and moves to one or other side-contact when current is applied to the coil.*

TMC TRADE MARK

Manufactured by the sole licensees:

TELEPHONE MANUFACTURING COMPANY LTD
HOLLINGSWORTH WORKS, DULWICH, LONDON SE21

Telephone: GIPsy Hill 2211

SPECIAL QUALITY PRODUCTION COMMERCIAL TYPE NUMBER (AMERICAN & MULLARD)		5636	5718	5840	5899	5902	6021	6205	5896
BRITISH SERVICES TYPE NUMBER		CV 3928	CV 3930	CV 3929	CV 477	CV 4029	CV 3986	—	CV 2698
STANDARD PRODUCTION COMMERCIAL TYPE NUMBER		EF 730	EC 71	EF 732	EF 731	EL 71	ECC 70	EF 734	—
DESCRIPTION		Short Suppressor Base RF Pentode	UHF Triode	RF Pentode	Variable-mu RF Pentode	AF Output Pentode	Double Triode	RF Pentode	Double Diode
HEATER	Vh (V)	6·3	6·3	6·3	6·3	6·3	6·3	6·3	Under development
	Ih (A)	0·15	0·15	0·15	0·15	0·45	0·3	0·15	
LIMITING VALUES	Va (V)	165	165	165	165	165	165	165	
	Vg2 (V)	155	—	155	155	155	—	155	
	pa (W)	0·55	0·9	0·8	0·75	3·7	0·7	0·8	
	pg2 (W)	0·45	—	0·35	0·35	0·4	—	0·35	
	Ik*(mA)	16·0	22·0	16·5	16·5	50	22·0	16·5	
*CAPACITANCES	cin (pF)	4·0	2·2	4·2	4·3	6·5	2·4	4·2	
	cout (pF)	3·4	0·7	3·4	3·4	7·5	†0·28 ††0·32	3·4	
	ca-gl (pF)	<0·02	1·45	<0·015	<0·015	0·2	1·5	<0·015	
TYPICAL CHARACTERISTICS	Va (V)	100	100	100	100	110	100	100	
	Vg2 (V)	100	—	100	100	110	—	100	
	Rk (ohms)	150	150	150	120	270	150	150	
	Ia (mA)	5·3	8·5	7·5	7·2	30·0	6·5	7·5	
	Ig2 (mA)	4·1	—	2·4	2·0	2·0	—	2·4	
	gm (mA/V)	3·2	5·8	5·0	4·5	4·0	5·4	5·0	
	μ	—	27	—	—	—	35	—	
	ra (kΩ)	110	4·7	260	260	15	6·5	260	
NOTES		At Ia = <100μA Vg3 = —8V approx.	Pout = 0·9W at f = 500 Mc/s.		At gm = 25μA/V Vgl = —14V approx.	Pout = 1·0W	Values are for each section except where stated.	This valve is the same as 5840 but with separate g3 connection	

ABRIDGED DATA

Notes †Section No. 1
††Section No. 2
*Capacitances are measured with external shield except for types 5718 and 6021.

✳ *Extremely Rugged*

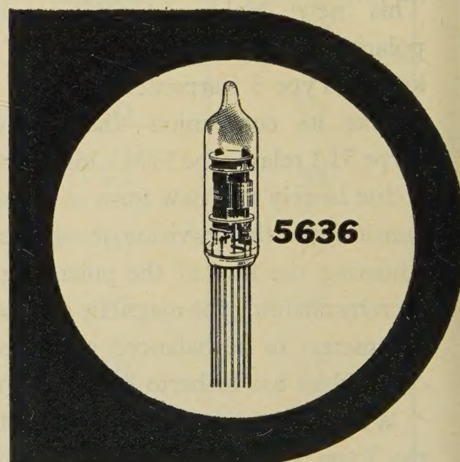
✳ *Shock Tested*

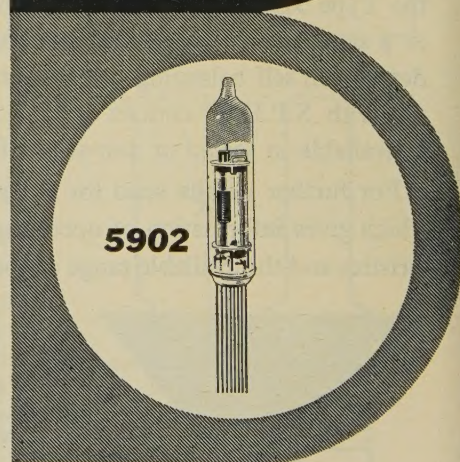
✳ *Vibration Fatigue Tested*

✳ *220°C. Life Tested*

✳ *Tested to U.S. Mil. Spec.*

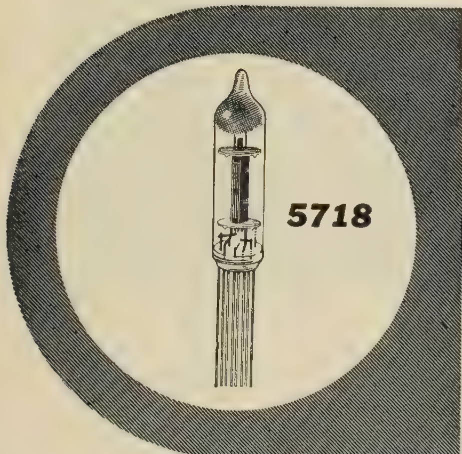
✳ *N.A.T.O. Preferred*

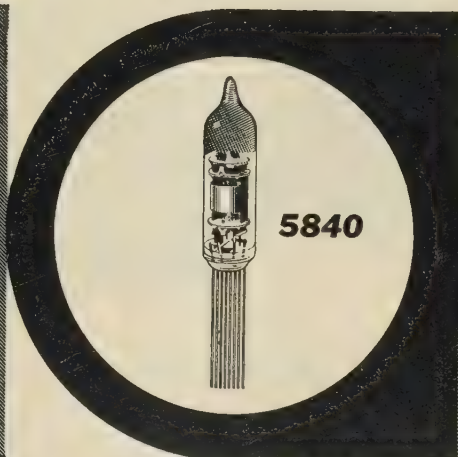


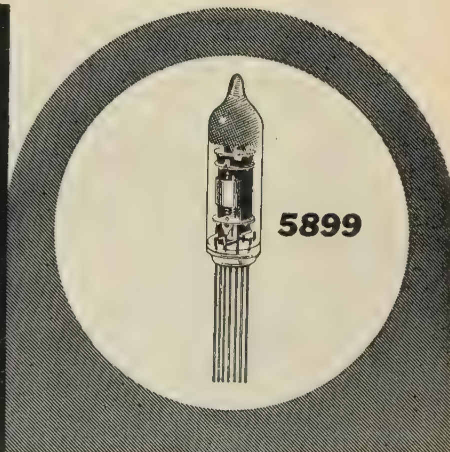


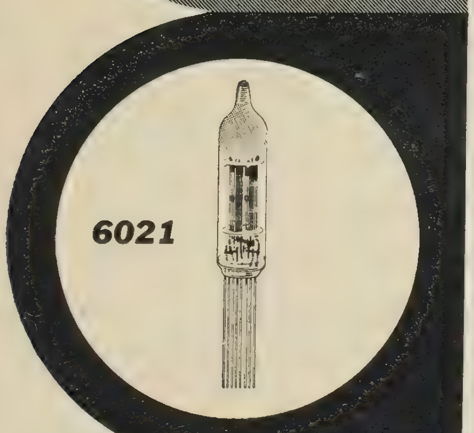
For the first time in this country, design engineers are offered a comprehensive range of internationally recognised, indirectly-heated subminiature valves. With many advantages in small size, weight saving and high mechanical and electrical reliability to Special Quality standards, these valves are widely used by the American and NATO forces.

Indirectly Heated Subminiature Valves Preferred Range

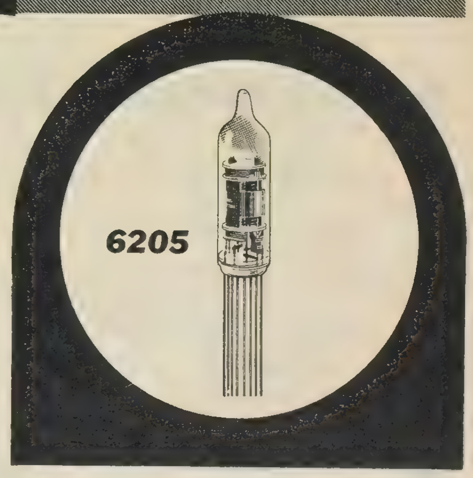








These Special Quality valves are now becoming available from Mullard under their American type numbers and are tested to U.S. MIL. specifications. The standard production versions are recommended for those applications where full MIL. assessment is not essential. In these applications certain electrical ratings may be increased and reference should be made to the relevant data sheets. These data sheets and information on all Mullard subminiature valves are freely available on request.



Mullard

Mullard

COMMUNICATIONS AND INDUSTRIAL VALVE DEPARTMENT

Mullard Ltd.
Mullard House, Torrington Place
London, W.C.1

ARMY WIRELESS SET TYPE C12

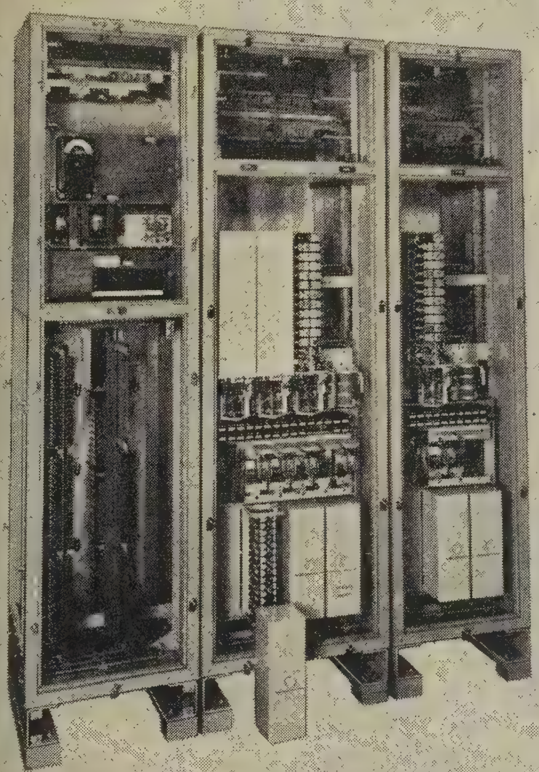
The Army Wireless Set Type C12 designed and manufactured by Pye Ltd., Cambridge to replace their world famous Wireless Set No. 19, has been adopted by the War Office for use in roles beyond the capabilities of the new VHF equipment.

Pye Ltd., specialize in the design and production of military mobile radio equipment. Their well-known wartime range of sets is continued in peacetime by the WS 62, still in quantity production for the War Office and many overseas Armies.

PYE TELECOMMUNICATIONS LTD · NEWMARKET ROAD · CAMBRIDGE

AUTOMATIC EXCHANGES

of facilities for Public and Private Service



4 + 24/8 + 49-line P.A.B.X. *One steel cubicle contains automatic equipment for 4 + 24-line exchange; two cubicles contain equipment for 8 + 49-line exchange. Incoming exchange calls are signalled on attendant's cabinet.*

50-line P.A.X. *One steel cubicle contains equipment for fifty extensions and six connecting circuits. Similar units can be added to extend the exchange to over 400 lines if required.*

50-line R.A.X. *Typical R.A.X. equipment showing distribution frame and automatic equipment. The exchange can be extended to 100 lines, if required.*

P.A.X. When a G.E.C. private automatic exchange handles the internal traffic, a P.B.X. operator is free to give proper attention to incoming and outgoing calls. Exchanges can be supplied ranging in size from ten lines to many hundreds.

by using G.E.C. Telephones

Telephone, Radio and Television Works, Coventry

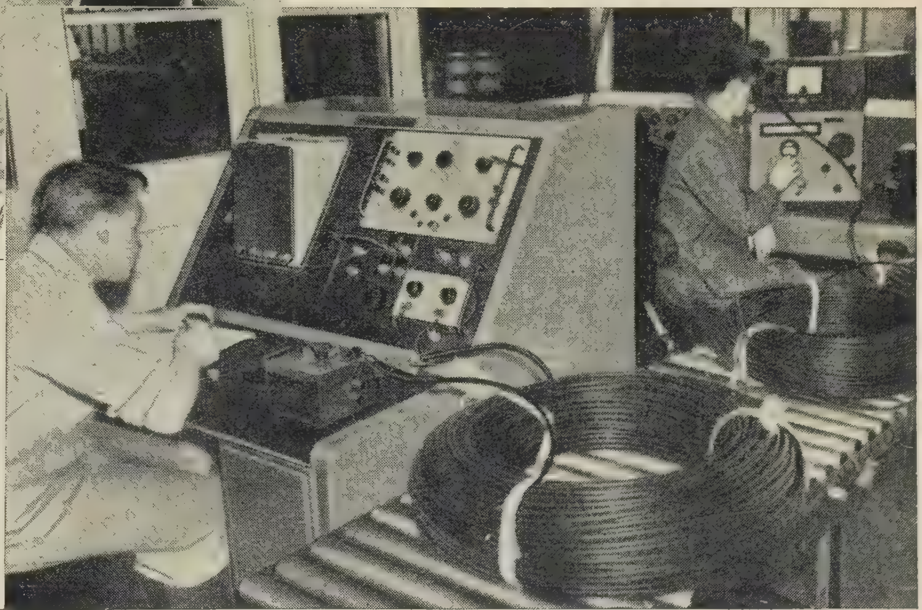
"Everything for Telecommunications by open-wire line, cable and radio; single and multi-circuit, and T.V. link; short, medium and long haul. Automatic and manual exchanges."

G.E.C.

Tested for endurance . . .

Every cable made by the fourteen members of the Cable Makers Association is rigorously tested for quality and endurance and has behind it an elaborate and continuous process of research. Together, the member-firms of the C.M.A. spend over one million pounds sterling every year on research and development —none of it wasted by duplication of effort, since one of the C.M.A.'s important functions is to ensure that essential knowledge, data and experience are freely and equally available to all members. The benefit to cable users of such co-operation is self-evident. It has helped to make Britain the largest exporter of cables in the world.

The Roman Warrior and the letters 'C.M.A.' are British Registered Certification Trade Marks



MEMBERS OF THE C.M.A.

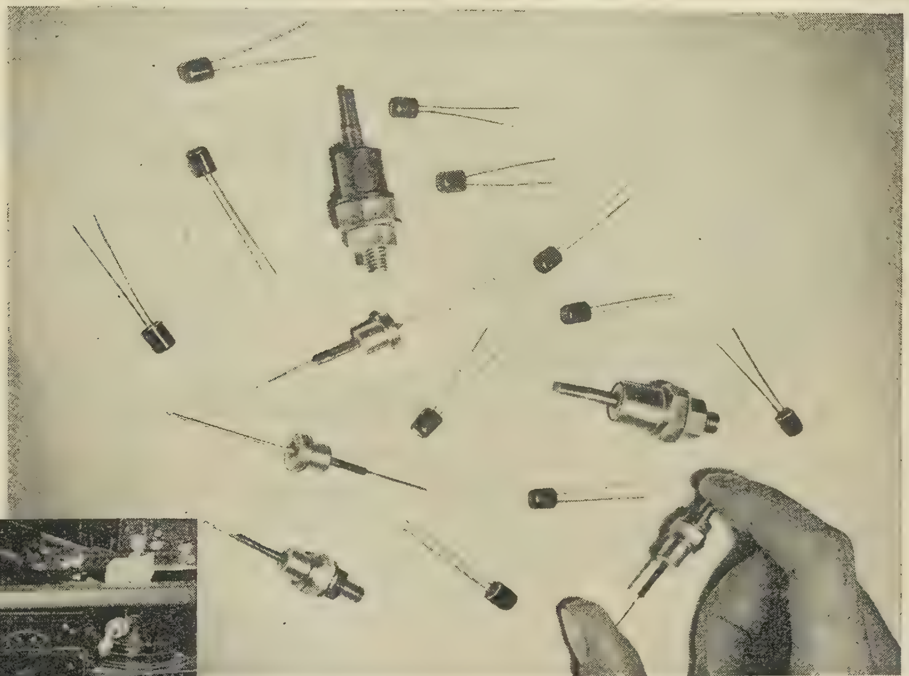
British Insulated Callender's Cables Ltd. Connollys (Blackley) Ltd · Enfield Cables Ltd. W. T. Glover & Co. Ltd · Greengate & Irwell Rubber Co. Ltd · W. T. Henley's Telegraph Works Co. Ltd · Johnson & Phillips Ltd. The Liverpool Electric Cable Co. Ltd. Metropolitan Electric Cable & Construction Co. Ltd · Pirelli-General Cable Works Ltd. (The General Electric Co. Ltd.) · St. Helens Cable & Rubber Co. Ltd · Siemens/Edison Swan Ltd · Standard Telephones & Cables Ltd. The Telegraph Construction & Maintenance Co. Ltd.

Insist on a cable with the C·M·A label

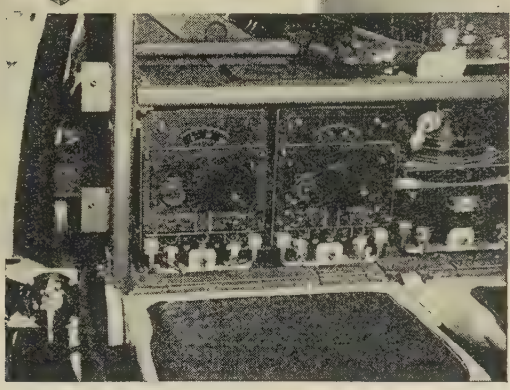
CABLE MAKERS ASSOCIATION, 52-54 HIGH HOLBORN, LONDON, W.C.1. Tel: Holborn 7633

CMA15

A wide range of applications
FERRANTI SILICON JUNCTION DIODES *and* SILICON POWER RECTIFIERS



Ferranti Silicon Junction Diodes are used in Smiths Flight System.



- LOW REVERSE CURRENT
- HIGH FORWARD SLOPE
- HIGH TEMPERATURE OPERATION
- SMALL PHYSICAL DIMENSIONS
- HIGH MECHANICAL STRENGTH

FERRANTI

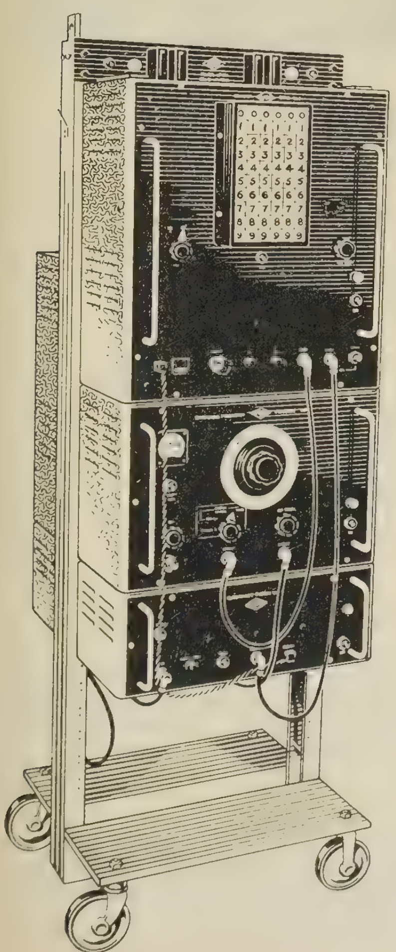
Ferranti Silicon Junction Diodes have been chosen for use in Smiths Flight System not only for their efficient operation, but also for their complete reliability, robust construction, small size and lightness in weight.

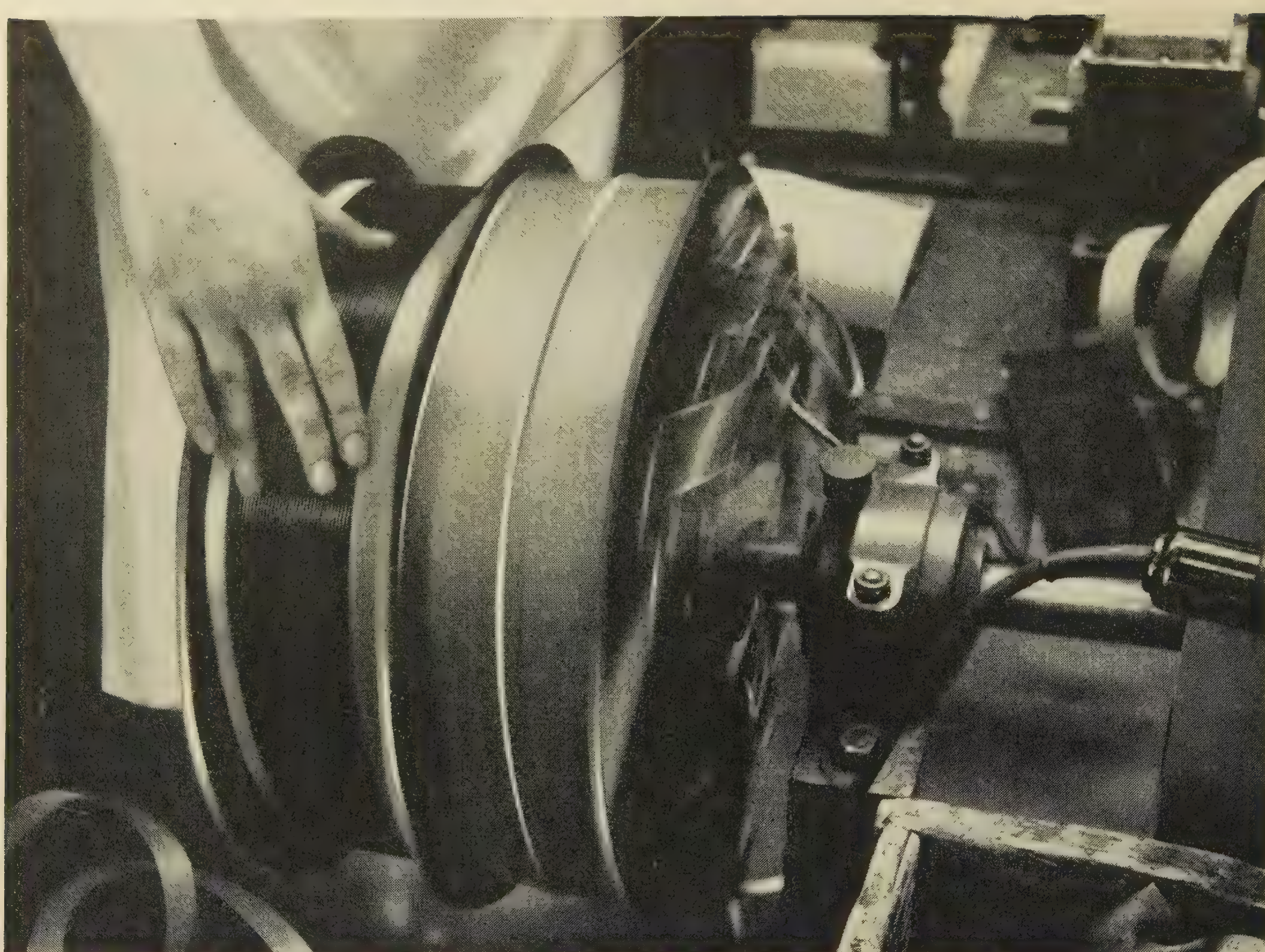
Ferranti Silicon Junction Diodes and Silicon Power Rectifiers have many applications in the aircraft, electronic, electrical and general engineering industries including aero engine controls, aircraft power supplies, radar systems, guided missiles, computers, indicating and recording instruments, process control and telephone equipment.

FERRANTI LTD · GEM MILL · CHADDERTON · OLDHAM · LANCS

London Office: KERN HOUSE · 36 KINGSWAY · W.C.2.

INDEX OF ADVERTISERS





THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 104. PART B. No. 16. JULY 1957

DISCUSSION ON

'FADING OF LONG-DISTANCE RADIO SIGNALS AND A COMPARISON OF SPACE- AND POLARIZATION-DIVERSITY RECEPTION IN THE 6–18 Mc/s RANGE'*

Mr. K. L. Rao (*India: communicated*): It is noticed that the assessment of the results has been made on the basis of distances. I suppose that cognizance may have to be taken of the direction from which these stations are received. The standard deviation in the diversity factor and the correlation coefficients appear to be generally more for the eastern stations than those observed from Schenectady. Listener observations here would appear to indicate that the fading of the eastern and western stations is in the nature of the records shown for Shepparton and Schenectady respectively. It may therefore be more difficult to combat fading of the type experienced from Shepparton in actual practice.

I suppose that no satisfactory system of diversity has been developed for the 'flutter' type of fading, even though lowering of receiver sensitivity and adjustment of the time-constant of the automatic gain control have some effect in reducing it. This type of fading appears to be prevalent for stations in the equatorial regions during periods of high sunspot activity and generally occurs during overlap of sunset periods between the transmitting and receiving points. This phenomenon appears to be peculiar to stations like Colombo, Singapore and Manila, no such fading being observed in transmissions from South Africa and Cairo. Sometimes flutter is observed from Australian transmissions also.

Messrs. G. L. Grisdale, J. G. Morris and **D. S. Palmer** (*in reply*): The results indicate that more distant stations give, in general, a lower fading speed and also a lower spatial structure size. An examination of the effect of direction would not be possible with the number of directions available in the experimental results. It is probable that direction would have some effect on the fading, particularly with respect to the paths passing through the magnetic pole regions. However, as recorded in Table 6, the diversity results show very little variation with either distance or frequency.

Mr. Rao's observation that the standard deviation of diversity factor is greater for eastern stations than for transatlantic stations has been checked by taking the mean standard deviation of results in Tables 1, 2, and 4 for both polarization and space diversity:

Station	Mean standard deviation of diversity factor	Number of results used
Shepparton	0·16	167
Colombo, Karachi, Delhi	0·28 (*a*)	479
Schenectady, Sackville ..	0·23 (*b*)	149
Moscow, Ankara ..	0·28	204

The difference between (*a*) and (*b*) is not very conspicuous, although it would be significant statistically, having regard to the large number of results used, so long as tests are assumed to be independent. As the tests were often made in groups, this assumption is not valid, and the significance becomes more doubtful.

No work was done on the problems of very rapid fading; our observations indicate that this is prevalent on the more distant transmissions.

* GRISDALE, G. L., MORRIS, J. G., and PALMER, D. S.: Paper No. 2239 R, January, 1957 (see 104 B, p. 39).

DISCUSSION ON

'COMPUTER INPUT AND OUTPUT, INCLUDING ANALOGUE-DIGITAL CONVERSION'

AT THE CONVENTION ON DIGITAL-COMPUTER TECHNIQUES, 12TH APRIL, 1956

Messrs. W. S. Elliott, R. C. Robbins and **D. S. Evans** (*in reply*):* Mr. Nettell said that in our digital servo we were converting to analogue with a 1 in 4000 accuracy to feed the servo motor. The details of the digital servo were not included in the paper, which would otherwise have been too long. We were in fact converting only the least significant 5 digits of the error to analogue and using that quantity to control the servo. Mr. Nettell referred to Dr. Faulkner's earlier paper, in which it is stated that a 1% accuracy is adequate in a good many situations in process control, and Mr. Nettell went on to speak of the possibility of simpler coded commutators using rubbing contacts. Such commutators are in fact available [Mr. Elliott showed a commutator of this type]. Mr. Nettell also referred to the possibility of eliminating the flying-spot scanner. In the two digitizers, each of 1 in 10000 accuracy, which we have on show here, there is in fact no flying-spot scanner. At the time when the work described in the paper was done (1948–51) there were not available to us the photo-sensitive devices now available.

Mr. Nettell suggested the use of codes based on the decimal system and said that decimal c.p. codes are now available. The digitizer here uses a decimal c.p. code [Mr. Elliott demonstrated the digitizer feeding a decimal numeric display]. We have also a similar digitizer, again with no flying-spot scanner, and a movement of 0·001 in on the shaft corresponds to 1 digit on the scale. The reading is decoded in electronic equipment, which as a matter of interest is made with printed-circuit transistor packages, and fed into the electric typewriter which types out the setting of the shaft position. [Mr. Elliott then gave a demonstration showing the accuracy of the apparatus; a feeler gauge inserted to move the shaft slightly caused the thickness of the gauge to be typed out.]

Mr. Nettell went on to speak of the problem of machine-tool control and said that in a linear commutator a digit spacing of less than 0·001 in would be required. In fact the counting method using a grating developed by Dr. Sayce and Mr. Guild, of the National Physical Laboratory, and applied to machine-tool drive by Mr. D. T. N. Williamson, gives that increased accuracy.

In reply to Mr. Coales, there would not seem to be an advantage in having any resetting facility at all, unless it were to the full accuracy of the system. The high redundancy of information on a linear-function-of-angle disc is admitted. There is, however, no need to economize in channel capacity in the circuits used to read the disc; if information is to be transmitted to a distance, redundancy can be reduced for that purpose. As we see it, resetting is an important advantage and would often justify the provision of the additional photo-transistors and circuits to read a fully coded commutator, as compared with the one photo-sensitive device needed with a counting method. In using the 'half-breed' system suggested by Mr. Coales there would be difficulty when the direction of turning of the shaft changed. Further, an advantage of the coded-disc method stated in the paper—and an essential part of the original requirement—is the ability to read out a function of shaft angle by suitable coding of the disc (Section 2.4 of the paper). This advantage would be lost in going to a counting or 'half-breed' method. The redundancy on a function-of-angle disc is not as great as on the linear-function disc.

Mr. Coales mentioned the difficulty of blemishes on the disc. In the paper we said that further work needed to be done on the photographic techniques. In fact in the discs demonstrated here, the mark length is much greater than the width and by integrating the light over the length of the mark we have been able to avoid any difficulty due to blemishes.

In reply to Mr. R. N. Jackson the scan rate was in fact high. If it were not high enough, the scan could easily be deflected in the direction of rotation of the shaft, and proportionally to its speed as indicated by a tacho-generator. This would be an interesting use of an analogue device to secure the correct functioning of a digital device. In parallel reading the scan-rate problem naturally does not apply. All the digits can be transferred at once to a register and can then be serialized while the shaft is moving on. The possibility of deriving a voltage analogue from the shaft position is mentioned in Section 1 of the paper, and we cannot agree with Mr. Jackson that this method seems preferable, for high accuracy. Mr. Jackson refers to applications of lower accuracy; we have mentioned lower-accuracy commutators in the reply to Mr. Nettell. We would go so far as to say that the use of a coded commutator, on a shaft carrying a potentiometer in a balancing circuit, is a good method of voltage-digital conversion. Mr. Jackson would require a voltage-digital convertor after obtaining the voltage analogue of his shaft. It would, of course, be absurd then to use a coded commutator for his conversion; the point is that for very high accuracies the coded disc is the only method of shaft-to-digital conversion and for lower accuracies it is the best.

In reply to Mr. Williams, we are staunch supporters of Nettell's principle, hence our rejection of geared shafts. We also agree with Mr. Williams in preparedness to depart from the strict principle by deriving the least significant digits in an analogue way, converting the analogue information to digital form and adding the resultant digits to those of higher significance obtained by direct coding. A quite specific example is given at Section 2.5 of the paper and incidentally is another example of mixed analogue-digital working, a subject mentioned by Mr. Dilger. Mr. Williams's suggestion of separating the noise signals produced from the index at one end of an interval from those produced from the index at the other end is most ingenious. We should be interested to learn how Mr. Williams proposes to effect this separation.

* This reply to the discussion was inadvertently omitted from *Proceedings I.E.E.*, 1956, 103 B, Suppl. No. 3, p. 449.

621.318.1

Paper No. 2136 M
Aug. 1956

AN EXPERIMENTAL STUDY OF HIGH-PERMEABILITY NICKEL–IRON ALLOYS

By C. E. RICHARDS, F.R.I.C., E. V. WALKER, B.Sc., and A. C. LYNCH, M.A., B.Sc., Associate Member.

(*The paper was first received* 6*th February, and in revised form* 1*st May*, 1956. *It was published in August*, 1956, *and was read before the* MEASUREMENT AND CONTROL SECTION 29*th January, and the* NORTH-EASTERN RADIO AND MEASUREMENT GROUP 4*th March*, 1957.)

SUMMARY

Because the properties of commercial alloys are inconsistent, designers of electrical equipment normally assume properties greatly inferior to those usually quoted. Further, very thin strips or sheets have losses greater than those predicted by simple theory. Factors which might influence either the ultimate performance or the consistency of the metals have therefore been studied, with particular reference to the 77/14/5/4 nickel–iron–copper–molybdenum alloy, using samples made by powder-metallurgy. If high permeabilities are to be attained the permissible tolerance on composition is very small, the limit for iron and nickel being only $\pm 0 \cdot 2\%$; further, any introduction of silicon makes the alloy susceptible to damage by water vapour in the annealing atmosphere. Using conventional melting techniques it is difficult to hold such tolerances and to avoid contamination by silicon.

The types of low-permeability layer which have been previously reported are not intrinsic and are probably due to impure metal which has been unsatisfactorily or incompletely annealed. Although metal having a permeability of 30 000 can be consistently made by strict control of composition and annealing, there is an unexpected component of a.c. loss which becomes serious at thicknesses less than about 30 microns.

Spiral cores, if free from strain, can have properties almost as good as those of the strip they are made from.

LIST OF SYMBOLS

a = Thickness of strip, cm.
f = Frequency, c/s.
g = Ratio of observed to calculated eddy-current loss.
k = A constant, $cm^{3/2}$.
Q = Ratio of reactance to resistance.
μ = Permeability, gauss/oersted.
ρ = Resistivity, C.G.S. units of resistance × cm.

(1) INTRODUCTION

The economical use of valuable raw material such as nickel is always important. The object of the work described in the paper is to economize in nickel by producing alloys of the Mumetal or Permalloy C type which are perhaps better and certainly more consistent than those now available. Designers must allow for the worst material they are likely to get, so that a consistent material can be used more economically than one which, though excellent when at its best, is liable to large variations from batch to batch.

The alloys in question have as their main constituents nickel (75–80%) and iron (12–16%), and as minor constituents molybdenum and often copper. Their initial permeabilities are quoted as 15 000–20 000 for Mumetal and 20 000–30 000 for 4–79 Permalloy.*[1,2] These data are presumably for commercial materials; other authors quote similar properties as typical. Designers, however, cannot use these figures, because some batches of material are much below this standard. Halsey[3] says: 'In the author's experience the initial permeability of thin Class A nickel–iron strip has been measured as low as 2 000 and as high as 40 000' (Class A means alloys of the type considered in this paper; Halsey had 50-micron strip cores in mind). Williams[4] makes almost exactly the same comment about 125-micron material delivered over a period of four years. These observations are, in the main, in line with those of the present authors. Material less than about 25 microns thick is even more troublesome. For example, two samples, 8 microns thick, obtained from entirely different sources had initial permeabilities of only 4 000.

The permeabilities commonly found are particularly disappointing, because certain alloys of substantially the same composition are claimed to reach initial permeabilities of over 100 000 if given a special, though not difficult, heat treatment.[5, 6] The only respect in which these alloys are claimed to differ from the normal ones is the absence of any metals more electropositive than manganese[7]—particularly magnesium.[8] Kersten,[9] working on a simplified model, shows that metals of low magnetostriction and anisotropy, such as those under consideration, should be relatively unaffected by traces of foreign elements in the lattice.

The problem of low permeability in these alloys has been attacked before. For example, Peterson and Wrathall[10] showed that by etching away the surface layers on soft iron, nickel–iron, and chromium–Permalloy, they could obtain metal of higher permeability and lower magnetic losses. This work has been confirmed by several investigators and extended, notably by Epelboin and his collaborators.[11] Hence an impression has grown that the production of a low-permeability high-loss surface layer is inherent in the rolling process (though neither Peterson and Wrathall nor Epelboin actually say so) and that the only way to get good-quality thin magnetic strip is to roll it relatively thick and to remove the surface layers by electropolishing or etching. Epelboin[12] patented the use of electropolishing for producing improved magnetic strip, though this appears to be scarcely a practical manufacturing process.

Previous work has not usually distinguished between trouble in the alloy itself and that caused by its insulation and by making it up into a spiral core. The present investigation consists mainly of the preparation and testing of small specimens free from insulating material; some separate tests, reported in Appendix 1, deal with the making of spiral cores. Much of the work was carried out on 50-micron material which, when this work was begun in 1946, was regarded commercially as fairly difficult to produce.

In judging these alloys, the authors have until recently confined their work to the low-field region and in so doing may have failed to help some prospective users of the material, notably the designers of magnetic amplifiers. Some of the conclusions in the paper, however, seem to be of quite general application and there is no reason why the materials which are satisfactory for low flux densities should not be the most useful at high flux densities as well.

* 4–79 Permalloy differs from Mumetal in containing no copper. These data are taken from well-known reference books, but the manufacturers do not necessarily agree with them.

The authors are at the Post Office Engineering Research Station.

Since this work began some manufacturers have, in collaboration with the authors, made changes so that their material is better and more consistent. For example, two manufacturers have, since January, 1952, been prepared to supply spiral tape cores, in a limited range of sizes, having the following guaranteed *core* permeabilities.[13]

0·004 in (100 micron)	10 000
0·002 in (50 micron)	7 500
0·001 in (25 micron)	5 000

measured at 50 c/s and $H = 0 \cdot 002$ oersted (r.m.s.).

Two manufacturers have, during 1955, begun to offer cores having permeabilities of 20 000 or more.

(2) DEFINITIONS

The following quantities are used in the paper to assess the quality of material:

μ is the initial permeability (the word 'initial' will be omitted), measured at 50 to 100 micro-oersted and at a low frequency.

g is the loss ratio, i.e. the ratio of the observed loss at low fields and low frequencies to the calculated eddy-current loss for a uniform strip. This is approximately equivalent to the η introduced by Feldtkeller[14] and the σ_a/σ introduced by Epelboin.[15]

From the user's point of view, the losses are measured by the quantity μQ, where Q is the ratio of the reactive and resistive components introduced by the core material;

$$\mu Q = \frac{3\rho}{2\pi^2 f a^2 g}$$

where f is the frequency, a is the thickness of the strip, and ρ is its resistivity measured with direct current.

The so-called a.c. resistivity is equal to ρ/g, but the term has been avoided in the paper because it may conceal the true mechanism of the excess losses.

In judging the consistency of a group of values of permeability, some allowance should be made for the likely scatter among any really high values. The use of values of reluctivity instead of permeability would be justifiable on every ground except familiarity; but when results are expressed in this way, the scatter of the values for the best materials is greatly reduced. For example, the range of reluctivities corresponding to permeabilities of from 15 000 to 20 000 is as great as for permeabilities from 40 000 to 120 000. Differences between observed values in the latter range should therefore not be regarded too seriously.

(3) EXPERIMENTAL TECHNIQUE

This Section describes the techniques in an advanced state of development. The reasons for using these techniques are not fully discussed, but some will be apparent from the results given in Section 4. Some of the results quoted later, however, were obtained with less well-developed methods which gave materials of lower permeability than was afterwards attained.

(3.1) Sources of Commercial Alloys

Samples of cold-rolled material were obtained, taken in general from various manufacturers' current production, although there were also a few ingots specially made and rolled in the factories. In either case batches of material were identified so that samples from the same ingot were available at various stages of the rolling processes. Some of these samples were subjected in the laboratory to further rolling of the kind described in Section 3.2.

(3.2) Precision Alloys

Alloys were also made in the laboratory by powder metallurgy. They are referred to in the paper as precision alloys, implying not only that the major constituents are in accurately controlled proportions, but that the alloys are unusually free from impurities. Unlike the commercial alloys they contain no manganese.

The raw materials were carbonyl nickel (A Grade), carbonyl iron (MP Grade), and commercial high-purity copper and molybdenum. The iron and nickel powders were freed from oxide and residual traces of carbon by long annealing in pure dry hydrogen at about 400° C. Batches of about 100 g were made up, providing compacts about $3 \times 1 \times 0 \cdot 3$ in. The powders were weighed to an accuracy of 1 mg, thoroughly mixed by grinding in an iron mortar and pestle or tumbling in glass vessels, and pressed (dry) at 25 tons/in^2 into compacts. The compacts were sintered for five hours in pure dry hydrogen at 1 350° C in recrystallized-alumina tubes, so that alloying took place by diffusion.

All rolling into strip was done cold, with rolls lubricated with a special rolling oil. After cold-rolling to $\frac{1}{8}$ in, a further heat-treatment at 1 050° C or higher was given. Below this thickness, after each reduction to half thickness the strip was annealed for about ten minutes in hydrogen at 800° C and gas-quenched.

(3.3) Preparation of Test Specimens

If the material is magnetically isotropic, as these alloys very nearly are, a punched ring is the ideal form of specimen. (The published information[16] shows considerable variation of properties in different directions, but it relates to an early alloy whose properties suggest that it was not of the precise composition now known to be needed.) Diameters down to $\frac{3}{8}$ in have been used successfully. The ratio of outer to inner diameter should not be too large, but for this low-field work a ratio of 3 : 2 has been found satisfactory. From very narrow strip, specimens were made by welding into a loop or by making up a spiral core.

For the experiments described in Section 4, the specimens were annealed in pure dry hydrogen at 1 050° C for three hours unless otherwise stated. The gas was prepared by passing cylinder (i.e. electrolytic) hydrogen through first a catalytic deoxidizer and then an activated alumina drier. After this treatment the dew-point of the gas was about −70° C and it contained so little oxygen that after passage through an alumina tube at 1 000° C its dew-point was not measurably altered. The dried gas was led to the furnace through metal or p.v.c. tubes, the only rubber in the system being in O-ring seals on the furnace tube. The furnace tube was of glazed silica or recrystallized alumina, the latter being essential for any work at temperatures above 1 100° C.

If the anneal was to be in wet hydrogen, the cylinder hydrogen was led through water in a gas-washing bottle and thence to the furnace tube.

Generally the specimens were suspended vertically from a rod of recrystallized alumina, but those thinner than 25 microns were laid in a molybdenum boat whose surface was covered with very pure alumina powder.

The specimens cooled at the natural rate of the furnace—nine hours from 1 050 to 200° C. The rate through the more critical range, 600–300° C, was about 85° per hour.

(3.4) Electrical Measurement

The specimens were placed in plastics formers, using either a single ring or a few rings stacked loosely with paper interleaving. A toroidal winding, usually of 20 turns of two wires twisted together beforehand, was wound on the former, taking care not

to distort the specimen. The result was a mutual inductor of about 100 μH. Coaxial instead of twisted wire has also been tried, and gives a very high coupling factor, but introduces an unidentified loss.

The inductance and loss were then measured with a Maxwell- or Owen-type bridge, using the device introduced by Wilde[17] to allow both source and detector to be earthed. Most measurements were made at 2 kc/s, but for the study of permeability at various depths by a technique already reported,[18] frequencies up to 10 Mc/s were used. The field strength, of the order of 100 micro-oersted, was kept so small that no change of inductance with field was detectable.

If the specimen was so thick that full penetration of the flux could not be assumed, a calculated correction was applied.

For the measurement of resistivity, a gap was cut in the ring specimen and its resistance was measured with a Kelvin bridge.

The thickness of the specimen was calculated from its weight, assuming for the density a value, $8 \cdot 73\ g/cm^3$, which had been determined on a number of annealed specimens.

(3.5) Metallography

Microstructures of the annealed precision and commercial alloys were studied. One or more of three types of section was cut, (*a*) parallel, (*b*) normal and (*c*) oblique to the rolling plane. Method (*c*) was applied only to very thin strip. The technique for polishing and etching the specimens is described elsewhere.[19] Precision alloys were found to etch much more slowly than commercial ones.

(4) EFFECT OF VARIATIONS IN PREPARATION

In all, four major variables—composition, sintering temperature, rolling, and annealing—were studied. They are discussed separately below, although to some extent their effects are interrelated.

(4.1) Composition

An examination was made of the effects of varying the proportions of the major elements, and of the effect of some trace impurities.

(4.1.1) Major Elements.

A comprehensive range of commercial alloys, whose magnetic performance had been carefully measured, was completely analysed; exceptional care was taken that the major constituents were determined as accurately as possible, and both nickel and iron were determined directly and not by difference. An extensive series of precision alloys was also prepared and the magnetic properties determined. Inspection of the results on the commercial alloys showed that copper, molybdenum, manganese, silicon and magnesium, considered independently, had no consistent effect on the properties of the metal. It was, however, possible to relate the iron content closely to the permeability. The precision alloys in which copper and molybdenum were carefully controlled at 5% and 4% respectively, and from which manganese was absent, followed the same pattern. The curves in Fig. 1 show the relationship clearly, and it is evident that for an alloy of the 5%-copper/4%-molybdenum type, free from manganese, and cooled at the usual rate after annealing, the iron content should be $14 \cdot 0 \pm 0 \cdot 2\%$.

(4.1.2) Trace Elements.

Carbon and nitrogen were known to have a profound effect on the magnetic properties of iron, and Schumacher showed[8] that, in a nickel–iron alloy, $0 \cdot 1\%$ of magnesium, present as magnesium oxide, could reduce the permeability to one-third of the very high value attainable in its absence. In the present work

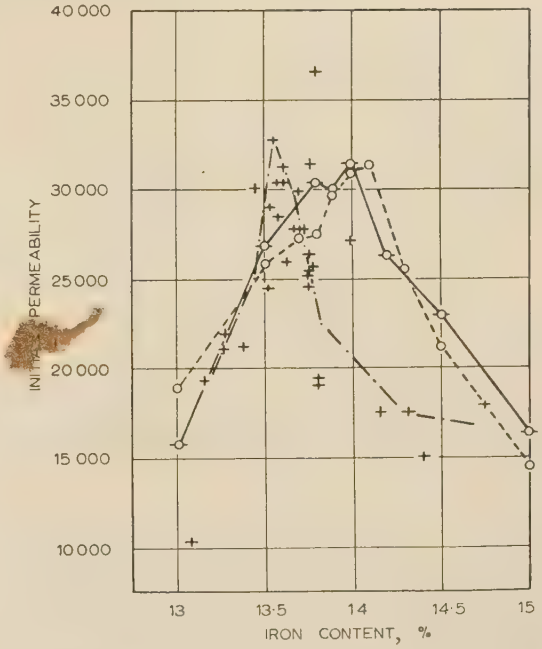


Fig. 1.—Variation of permeability with composition.

—·— + Commercial alloys.
——O—, ----O Precision alloys from two sources.

no attempt has been made to study the effect of these particular contaminants. The high purity of the raw materials and the techniques adopted have ensured the virtual absence of carbon,* nitrogen, aluminium, magnesium and manganese.

Silicon only has been studied; its effects are considered in detail in Appendix 9.2 and only those more directly concerning the electrical engineer are summarized here. Silicon and its compounds are likely to be present in the raw materials and may also be picked up during mixing and heat treatment. Silicon contamination from furnace linings is possible even if these are of alumina and contain silica only as an impurity, while in melted alloys pick-up from the crucible is almost inevitable. These effects are due to the instability of silica and silicates under reducing conditions at high temperatures, the reactivity of elementary silicon and oxygen, and the existence of a volatile silicon monoxide. Silicon reaching the alloy as monoxide or as the element can dissolve in it and disperse; in this state it will still be susceptible to the action of water vapour or oxygen at high temperatures and some of it will re-oxidize to silica. This silica, being insoluble in the metal, will precipitate and result in the annealed metal cooling in a state of strain. The rate of breakdown of silica increases very rapidly with increasing temperature, so alloy which has been exposed in any way to silica contamination will be more seriously affected by high-temperature (1 300° C) treatment than by low (1 050° C). Alloy which has been melted (>1 400° C) will be more affected than even that which has had a high-temperature sinter. Contamination during the annealing cycle can be suppressed by the use of a protective atmosphere of moist hydrogen, although cooling below about 1 100° C should be in dry hydrogen if there is already an appreciable amount of silicon in the alloy.

The continuing improvement in the permeabilities of silicon-free alloys during long anneals in alumina tubes (see Table 4) cannot be attained when annealing is done in silica tubes, even at 1 050° C; Table 1 gives a typical example.

* Analyses, kindly carried out by Mr. Still of the Research Laboratories of the General Electric Company, have shown that the carbon content is in the region of $0 \cdot 005$ to $0 \cdot 01\%$.

Table 1

PERMEABILITY AFTER PROLONGED ANNEALING IN ALUMINA OR SILICA TUBES

Annealing periods	Permeability	
	After annealing in alumina tube	After annealing in silica tube
hours		
3	33 000	—
9	42 000	—
17	48 000	50 000
34	48 000	46 000
52	47 000	39 000

At each of the periods mentioned, a few of the specimens were removed for measurement, and the rest left for further heat-treatment.

(4.2) Sintering

It is possible to sinter compacted powders at temperatures from about 1 000° C upwards, but longer soaking is necessary at the lower temperatures if similar results are to be attained. Table 2 records the permeabilities of strip rolled from compacts which, though otherwise similar, had been sintered for five hours at different temperatures.

Table 2

EFFECT OF SINTERING TEMPERATURE ON PERMEABILITY

Sintering temperature	Permeability at various thicknesses			
	200 microns	100 microns	50 microns	30 microns
°C				
1 150	19 500	23 000	30 000	35 500
1 250	26 000		30 000	33 000
1 350	34 500	33 000	40 500	41 500

A secondary effect of sintering temperature, in altering the effect of silicon, is discussed in Section 4.1.2.

(4.3) Rolling

It was thought possible that the diameter of the rolls used for cold reduction might be responsible for differences between reputedly similar metals. To check this point, comparative tests were made, in which commercial ingots were rolled to 200 microns in the usual way, and divided into two parts which were reduced to 50 microns on different-sized rolls. The 50-micron strip was annealed and tested in the usual way. Table 3 gives the results.

The last two entries refer to the same ingot, but the thickness for the last one was 106 microns.

There is no indication that varying the roll diameter between 8 in and 1 in makes any significant difference to the product.

Table 3

EFFECT OF ROLL DIAMETER ON PERMEABILITY AND LOSS RATIO

	8 in rolls		3⅛ in rolls		2⅜ in rolls		1 in rolls	
	μ	g	μ	g	μ	g	μ	g
Six different alloys	29 500	1·53			30 000	1·33		
	16 000	1·34			17 500	1·23		
	25 500	1·46			24 500	1·38		
	25 000	1·35			28 000	1·25		
	32 500	1·36			36 500	1·28		
			24 500	1·37			24 500	1·35
	33 000	1·33					30 500	1·23

(4.4) Heat Treatment

All nickel–iron alloys of the type here considered are subjected to two kinds of heat treatment: intermediate, at given stages during cold rolling; and final, after all rolling, insulation and other manipulation has been done to the metal. These heat-treatments are generally described as annealing, and for convenience this term is used in the paper, but, as will be seen below, annealing in the true metallurgical sense is only part of the story.

The possible variants in the annealing process are time, temperature and atmosphere; with any two of these fixed, the significance of the third can be determined.

(4.4.1) Duration of Final Anneal.

The 3-hour annealing period used for most of this work was chosen, when the work was begun, because some commercial 50-micron alloys then available reached a higher permeability after a 2-hour anneal than after 1 hour; the 3-hour period was therefore thought to give a suitable margin. Later experiments showed, however, that this period was not always adequate. Table 4 shows the improvement in overall permeability obtained in precision alloys by increasing the annealing time, and the lower line in Fig. 2 shows the state that may be reached with a moderately thick specimen with too short an anneal, in which the outer part of the strip has reached nearly its best properties while the middle part is still of comparatively low permeability. If this effect is combined with that described in Fig. 4, the specimen with a 3-hour anneal is like one reported by Epelboin and Marais[11] which, however, they did not try to account for. The observed increase in losses is associated with an increase in crystal size.

The 3-hour period has nevertheless been retained for experimental work, as it seems to be reasonably useful for strips of 50 microns and less. Any apparently poor results for thicker strips should be regarded with caution.

(4.4.2) Final Annealing Temperature.

Preliminary experiments showed that a 3-hour final anneal at temperatures below 900° C led to only low permeabilities—about 1 700 at 500° C and 7 000 at 700° C. Commercial and precision alloys, 50 microns thick, annealed at temperatures between 950° C and 1 300° C gave the results shown in Fig. 3. Typically the

Table 4

VARIATION OF PERMEABILITY AND LOSS RATIO WITH PERIOD OF ANNEALING

Time of anneal at 1 050° C	Precision alloy, 43 microns thick			Commercial alloy, 57 microns thick	
	μ	g	g to be expected if thickness had been 57 microns*	μ	g
hours					
1	43 500	1·38	1·27	35 000	1·40
1 + 3	56 000	1·46	1·35	41 000	1·45
1 + 3 + 4	66 000	1·51	1·40	45 000	1·51
1 + 3 + 4 + 8	72 000	1·60	1·49	43 000	1·57
1 + 3 + 4 + 8 + 8	77 000	1·63	1·52	41 000	1·57
1 + 3 + 4 + 8 + 8 + 8	72 000	1·69	1·58	46 000	1·57
64	74 000	1·68	1·57	44 000	1·62

* See Section 5.2.

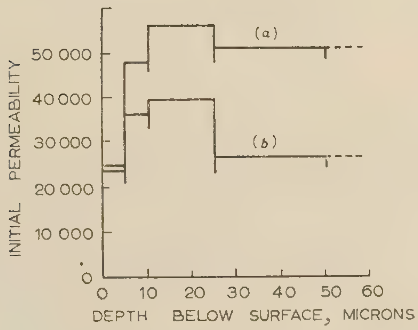


Fig. 2.—Permeability at various depths below the surface in a 200-micron strip.

(a) Annealed for 3 + 12 hours.
(b) Annealed for 3 hours.

The values are averages for arbitrarily-chosen ranges of depths; it is not suggested that there are really any discontinuities in permeability.

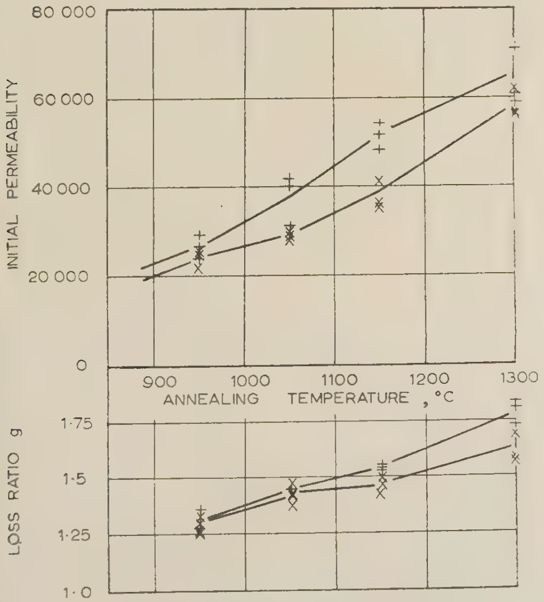


Fig. 3.—Permeability and loss ratio after annealing at various temperatures.

+ Precision alloys.
× Commercial alloys.

permeability of all the metals increases as the annealing temperature rises; the loss ratio g also increases, again as the crystal size increases. Provided that the annealing temperature is high enough, the result appears to be a function of both time and temperature; for example, a very long anneal of a precision alloy at 1 050° C gives results similar to a short one at 1 300° C.

(4.4.3) **Final Annealing Atmosphere.**

Two types of atmosphere are considered; those based on substantially pure hydrogen, in which the important impurities are water and oxygen (which is converted to water in the furnace), and those based on cracked ammonia.

(a) *Purity.*

Although pure dry hydrogen is usually the best annealing medium the necessary high purity is not always attainable in commercial work; cylinder hydrogen contains varying small amounts of oxygen and water vapour. The effect of these impurities was studied using samples annealed in hydrogen saturated with water vapour at room temperature.

Melted alloys annealed in wet hydrogen never reach as high a permeability as those annealed in the dry gas. With sintered alloys, as already mentioned, there is less difference. Table 5 summarizes a number of experiments.

Table 5

EFFECT OF WET ANNEALING GAS ON PERMEABILITY

Thickness of metal, microns	50	30	15	8
	Ratio $\frac{\mu \text{ after wet anneal}}{\mu \text{ after dry anneal}}$			
Melted alloys ..	0·63	0·56	0·50	0·47
Sintered alloys ..	1·04	0·98	0·75	0·78

This effect was examined in more detail by measuring the permeability at different depths; the results are shown in the upper half of Fig. 4. Thus the combination of the melted alloy and wet gas produces a low-permeability layer, of depth about 10 microns, on each surface of the material; yet either condition alone is permissible, the melted alloys being nearly free from this effect if annealed in dry gas, while wet gas has little effect on sintered alloys, except when the strip is less than about 30 microns thick. These results may be compared with those reported by earlier workers—e.g. the results [10, 11, 20] in the second row of

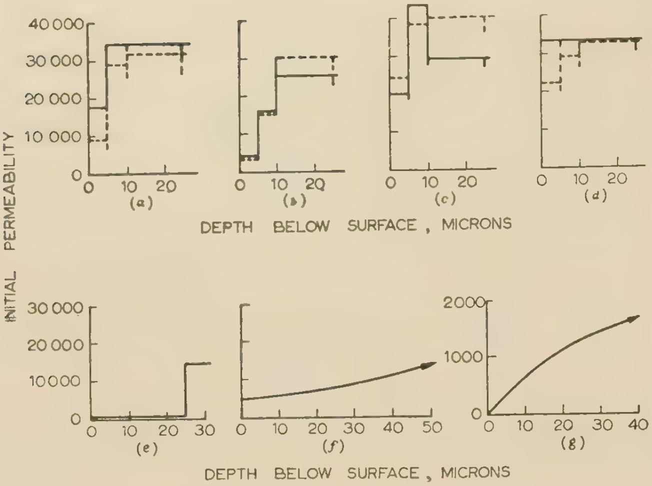


Fig. 4.—Permeability at various depths below the surface.

(*a*) Dry gas, commercial alloy.
(*b*) Wet gas, commercial alloy.
(*c*) Dry gas, laboratory alloy.
(*d*) Wet gas, laboratory alloy.
(*e*) Peterson and Wrathall. 350-micron strip.
(*f*) Epelboin and Marais. 500-micron strip.
(*g*) Feldtkeller, using a 36%-nickel alloy. 335-micron strip.
———— 50-micron strip. — — — 200-micron strip.

The values in the upper row of graphs are, as in Fig. 2, averages for arbitrarily-chosen ranges of depths.

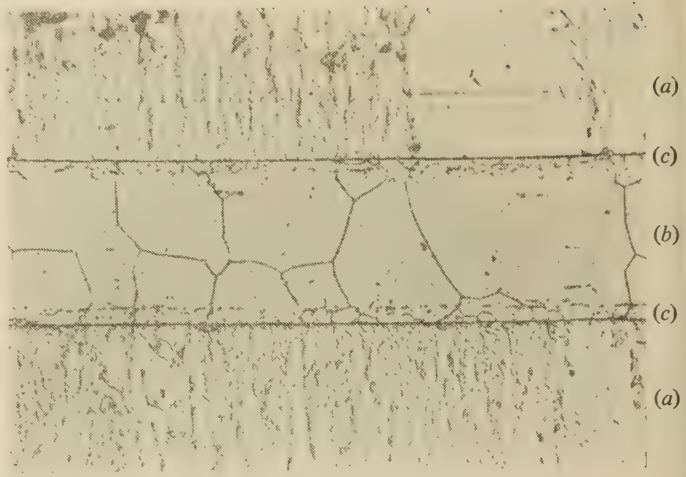


Fig. 5.—Surface layers on commercial alloy.

Magnification, ×400.
(*a*) Protective nickel plating.
(*b*) Specimen.
(*c*) Surface layer.

Fig. 4, which resemble the second diagram in the upper row rather than the first.

Further evidence was obtained by metallography. In 50-micron strip annealed in dry hydrogen the characteristic grain diameter, measured parallel to the rolling plane, was 45–50 microns for both kinds of strip; this fell with the thickness of the strip, reaching about 20–25 microns in 8-micron material. Precision alloys after wet annealing were indistinguishable from those annealed dry, but commercial alloys had grain diameter about 10% smaller. The grain diameters of 8-micron precision and commercial alloys were respectively 40% and 20% smaller when wet-annealed.

Precision alloys were much freer from inclusions than commercial alloys. Commercial alloys, if annealed in wet hydrogen, always had surface layers of small crystals and oxide particles, and sometimes showed them even if annealed in pure dry hydrogen. The appearance of these surface layers is shown in Fig. 5. Precision alloys, whether annealed in dry or wet hydrogen, never developed these layers. Some commercial strip rolled by the authors to 8 microns failed to develop surface layers on wet annealing; no explanation is offered for this anomalous result.

During a wet anneal, commercial alloys, as well as forming valleys at the grain boundaries, developed a spottiness on the crystal faces; this effect is visible to the naked eye since it causes the annealed commercial alloys to have a misty appearance. Sintered alloys did not show this effect.

The above results refer to the effect of the final anneals only, intermediate anneals having been done in pure dry hydrogen. Tests were, however, made on commercial alloys to find the effect of impure gas in the intermediate anneals when pure dry gas was used for the final anneals; Table 6 gives the results.

Table 6

EFFECT OF WET GAS IN INTERMEDIATE ANNEALS

	μ after intermediate anneals in dry hydrogen	μ after intermediate anneals in wet hydrogen
3 different commercial alloys 50 microns thick	29 700	26 300
	15 500	16 200
	22 700	25 000

Thus a moist atmosphere during the intermediate anneals seems to have no deleterious effect.

(*b*) *Use of Atmospheres of Nitrogen and Hydrogen.*

Cracked ammonia is often used commercially instead of hydrogen. It is essentially a mixture of one volume of nitrogen to three of hydrogen, and is exceedingly dry. It may, however, contain traces of ammonia, and even if this is washed out of the supply, some more will be formed during the passage of the nitrogen–hydrogen mixture through the furnace. The equilibrium mixture of gases will contain less ammonia at high temperatures than at low ones, so there are two basic conditions to study—those in which the ammonia content of the gases increases during its passage through the furnace and vice versa.

Four sets of samples were annealed respectively in (*a*) pure hydrogen, (*b*) 25% nitrogen, 75% hydrogen, (*c*) 50% nitrogen, 50% hydrogen and (*d*) 25% nitrogen, 75% hydrogen, containing enough ammonia to ensure that cracking took place in the furnace.

The results in Table 7 show that the presence of nitrogen with or without ammonia has no noticeable effect on the specimen.

(4.4.4) **Cooling Cycle.**

When the alloy is cooled after the annealing treatment, an important range is from about 600° C to 300° C, in which the furnace usually cooled at its natural rate of about 85° C per hour.

It has been shown by Boothby and Bozorth[5] and by Assmus and Pfeifer[6] that, with certain alloys, permeabilities over 100 000 can be reached either by cooling at a controlled rate or by quenching from above 600° C followed by baking at a temperature between 450° C and 500° C for a short time. The 77/14/5/4 alloy behaves in this way. The authors have found that both commercial and precision alloys can be so purified by heating at 1 300° C in hydrogen that they will develop permeabilities of 70 000 to 80 000 by furnace cooling at the above rate. But in precision alloys a further increase to over 100 000 can be obtained by quenching from 600° C and subsequently baking for 1½ to 2 hours at 460–470° C. This work is not com-

Table 7

EFFECT OF NITROGEN–HYDROGEN MIXTURES ON PERMEABILITY AND LOSSES

	Hydrogen		25% nitrogen, 75% hydrogen		50% nitrogen, 50% hydrogen		25% nitrogen, 75% hydrogen, and trace of ammonia	
	μ	g	μ	g	μ	g	μ	g
Commercial alloy ..	27 500		28 500		30 000		28 500	
3 high-purity alloys	25000		24 000	1·33	23 000	1·31	26 000	1·34
	31 500	1·21	31 000	1·21	31 500	1·19	29 500	1·24
	40 000	1·45	43 000	1·37	39 500	1·38	45 000	1·42

plete, but the findings confirm those of the previous workers quoted above.

(5) PROPERTIES UNAFFECTED BY VARIATIONS IN PREPARATION

Certain unexpected properties are found in both commercial and precision alloys.

(5.1) Variation of Permeability with Thickness

A large fall in permeability is usually thought to be unavoidable when the thickness of strip is reduced below about 100 microns; and if thin strip normally carried surface layers of the depth previously reported, then its mean permeability would necessarily be low. Microscopic examination of suitably-prepared specimens has shown the presence of shallow grooves at the grain boundaries and, in commercial alloys, of surface layers of inclusions and abnormally small crystals. These grooves are formed by thermal etching.[21] This is the revelation of microstructure of the surface of a metal after it has been heat-treated in a reducing or neutral atmosphere. The effect is caused by the migration of metallic ions from the grain boundaries which results in grooves being formed there. The depth and width of these grooves are of the order of 0·5–1 micron and 1–2 microns, respectively; there is no way of preventing their formation.

If materials are annealed as described in the paper, however, permeabilities can be obtained which, though somewhat lower than those of thicker strip, as shown in Fig. 6, are still of the same order.

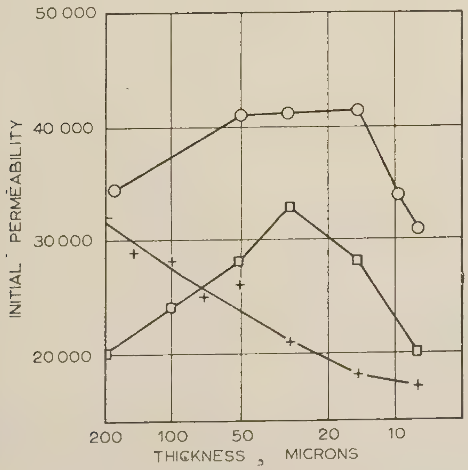


Fig. 6.—Change in permeability as material is rolled to smaller thicknesses.

○ Sintered at 1 350° C.
⊡ Sintered at 1 200° C.
+ Commercial (melted).

(5.2) Excess Losses in Very Thin Strip

For low frequencies the eddy-current loss in laminations is calculable, but the observed loss is usually somewhat greater. This effect is qualitatively well known in both nickel–iron and silicon–iron. Most of this extra loss varies with frequency and field in the same way as does eddy-current loss, and it can therefore be expressed by a ratio, g, of observed loss to calculated eddy-current loss, where g is nearly independent of frequency.

This ratio g is found to be high in thin strip. The line in Fig. 7

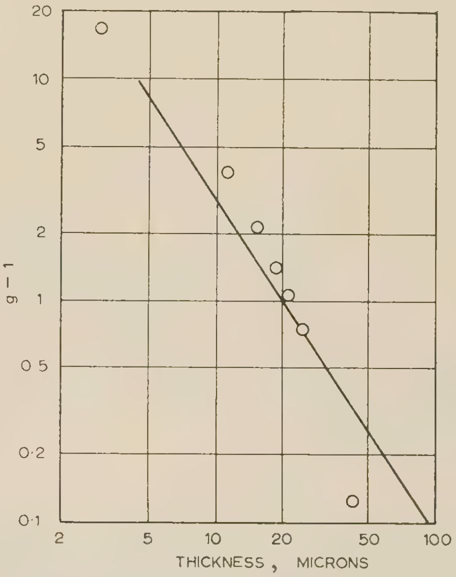


Fig. 7.—Variation with thickness of loss ratio g.

The line represents the mean of about 40 of the authors' results; the points are those reported by Abgrall and Epelboin.[22]

shows the mean of about 40 of the authors' results, for both commercial and precision alloys; the seven points are values reported by Abgrall and Epelboin.[22] The line is represented quite well by the empirical equation

$$g - 1 = ka^{-3/2}$$

where $k = 90$ if the thickness a is in microns. Epelboin says that this effect is independent of the method by which the thin strip is produced, whether by rolling or electropolishing. He interprets it as due to shape anisotropy. A more detailed way of expressing this is that, since the anisotropy of these alloys is negligibly small, there is no reason for a domain structure to exist, except perhaps near the surface, where something analogous to a system of closure domains is likely to form. The loss in a domain-and-wall structure is easily shown to be greater than

that in the uniform material considered when calculating the eddy-current loss.[23]

This anisotropy existing only near the surface may also be responsible for some anomalous values of the temperature coefficient of permeability which have been observed in thin strips.

An alternative explanation of the excess loss is that it arises from damping of the relaxation of the electronic spins.[24] An estimate[25] of the magnitude of this damping makes it about one-tenth of the observed effect; but this estimate is necessarily very rough, as some of the necessary data, e.g. the exchange energy, are not available for these materials. On this theory, g should be constant at all low frequencies (which it very nearly is) and should be proportional to a^{-2}, not $a^{-3/2}$. The results now reported can be reconciled with this theory only by supposing that there is yet another source of loss (e.g. the domain-and-wall structure) present as well.

(5.3) Effect of Mechanical Strain

All alloys of this type are strain-sensitive, i.e. their properties are changed by mechanical strain and in particular their permeability is lowered. The effect is reversible for small strains but irreversible for large.

Elastic distortion can reduce the permeability to about half its normal value and increase the loss ratio g by perhaps 0·5. When the applied force is released, the permeability and loss return immediately to values near their original ones, though complete recovery (like complete mechanical recovery) may take a few hours.

But these materials are so soft that quite small forces may do permanent mechanical damage. If so, the permeability falls to even lower values, and it cannot be restored even by re-annealing. Annealing is fully effective only if the previous cold working has been severe, as it is in cold rolling; ordinary mechanical damage, although it may be locally severe, leaves most of the material only lightly worked, and these areas will not recrystallize when annealing is attempted. The following experiment illustrates the effect:

Permeability of an annealed 50-micron ring ..	37 000
Permeability after the ring was folded round a pencil and then flattened	4 000
Permeability after re-annealing	20 000

(6) DISCUSSION

The results set out in the paper show that, if certain conditions are fulfilled, there is no difficulty in consistently making nickel–iron alloys which are capable of developing very high initial permeabilities.

The first and most important condition is that the composition must be held within limits of the order of ±0·2%; the second is that the final anneal shall not permit the formation of oxides within the metal. The resultant metal should then have a permeability of over 30 000 at thicknesses of 50 microns (0·002 in) or more, and 25 000 at 12 microns.

The first condition arises from the need to produce an alloy having simultaneously low magnetostriction and low anisotropy. The closeness of control now suggested has not previously been thought either necessary or possible; analyses of commercial alloys show a much wider spread, and it is unlikely that melted alloys can be kept always within so narrow a range.

Both the magnetostriction and the anisotropy are temperature-dependent; and as the permeability is roughly proportional to their reciprocals, it varies abruptly with temperature as they approach zero. The temperatures at which these quantities pass through zero depend critically on the composition, and so, therefore, does the temperature coefficient of permeability. Data for the temperature coefficient in the range 20–65° C have already been published;[26] they relate to alloys containing 5% copper and 4% molybdenum, and show a variation between almost zero and 0·8%/°C for the range of compositions liable to be used commercially. To control the coefficient within the limits 0·2 to 0·3%/°C, the iron content of the alloy must be kept between 14·0 and 14·1%.

The second of the conditions laid down above may require the use of very pure dry hydrogen during the annealing cycle. With commercial alloys which contain significant amounts of such reactive metals as manganese, magnesium and silicon, the dew-point of the effluent furnace gases should be no higher than −60° C. Precision alloys, such as have been described, are substantially free from these metals in any reactive form and may be annealed in commercial hydrogen. There seems no reason to expect worse results if cracked ammonia is used, provided it is as free from oxygen and water vapour as the hydrogen should be.

It is believed that, in an alloy of the type under consideration, heat treatment in hydrogen is not a simple bright anneal. It allows not only recrystallization, which is fairly rapid, but it also removes impurities by a diffusion process which at 1 050° C takes several hours. The hydrogen must have a long enough contact with the metal to diffuse right into it and ‘clean it up’, mainly by reducing oxides to metal. It has been shown that the presence of silica in a metal which is never heated beyond 1 050° C is relatively harmless. In a melted metal, which must have been raised above 1 400° C, or in one sintered at a high temperature for a long time, some silica can break down to silicon and dissolve in the alloy. If there is enough water vapour in the furnace atmosphere during the final anneal, it can give rise to oxygen which diffuses into the metal, oxidizes some of the silicon, and leaves a metal which is permanently strained.

This final annealing of impure metals in impure gas is believed to be the cause of the surface layers, of the order of 10 microns thick, regarded by some authors as responsible for poor performance. These layers are not unavoidable; they are due to the co-existence in the annealing furnace of water vapour and a reactive substance—usually silicon or magnesium—in the metal. In all but the thinnest metal they can be reduced to negligible proportions by proper annealing, and they are not formed on precision alloys. It seems likely that all the previously-reported layers were of this type. Epelboin and Marais's finding[11] that the thickness of the surface layers is independent of the thickness of the strip, provided that the heat-treatment remains the same, is consistent with this view; it is what would be expected for layers formed by the action of the annealing gas for a limited time.

With commercial alloys annealed at 1 050° C there is little to gain from annealing times exceeding three hours; precision alloys, however, may continue to improve for 20 hours. An annealing temperature of 1 150° C is better than 1 050° C, but unless the atmosphere can be very closely controlled a higher annealing temperature will damage commercial metal.

Within the limits of the experiments attempted the cold-rolling procedure did not seem to change the properties of the metal seriously, though the roll diameters used ranged from 8 in to 1 in. Reduction between anneals should be by about 50%, but the figure is not critical; overrolling is perhaps preferable to underrolling.

Although strip can be rolled very thin without undue loss in permeability, magnetic losses increase sharply with reduction in thickness, and with strip thinner than 25 microns they become important. For example, when simple calculation shows that 17-micron strip is thin enough, in fact 7-micron metal should be used. This may mean a call for even thinner metal than has yet been made. The cause of this additional loss is not yet known.

Although the properties in low fields have been taken as the criteria in all the work reported here, there is no reason to think that the processes developed are in any way peculiar to such applications. The high-field properties of these alloys, so far as they have been checked, are also good.

It is pertinent to ask how far this laboratory work can be applied industrially. A considerable improvement appears to be possible in the properties of alloys made by the conventional melting process by attention to (*a*) precision of composition and (*b*) more careful final annealing. It may, however, be impracticable to make precision alloys on a factory scale by melting, and inferior performance may therefore be unavoidable in alloys made on a large scale. The biggest demands from industry must continue to be met with this type of material, for the extra cost of precision alloys rules them out for many purposes.

For users of very thin strip, say 50 microns and under, the situation is a little different; here the cost of the final product is mainly that of rolling, so that the use of expensive raw materials, such as metal powders, is justified by even a small improvement in the finished product. A useful improvement has in fact been achieved: alloys have been produced in factory conditions which are nearly as good, and just as consistent, as the laboratory products described in the paper. For airborne and military uses, in which size and weight are of paramount importance, these precision alloys offer considerable advantage. A core known to have a permeability of 15 000 can be smaller than one which might have only 5 000, a very usual design figure, and needs less than one-third of the copper for winding.

(7) ACKNOWLEDGMENTS

The information used in the paper is published by permission of the Engineer-in-Chief of the Post Office. The authors are indebted to their colleagues at the Post Office Research Station for the help they have given with the very large programme of experimental work, and to the manufacturers who have supplied samples for test.

(8) REFERENCES

(1) BOZORTH, R. M.: 'Ferromagnetism' (Van Nostrand, New York, 1951), p. 159.
(2) BRAILSFORD, F.: 'Magnetic Materials' (Methuen, London, 1948), p. 125.
(3) HALSEY, R. J.: 'Laminated Ferromagnetic Cores at Very Low Inductions for Use in Line Communication', *Proceedings I.E.E.*, Paper No. 890 M, October, 1949 (**97**, Part II, p. 141).
(4) WILLIAMS, M. O.: 'The Assessment of Core Material Variability', *A.T.E. Journal*, 1953, **9**, p. 121.
(5) BOOTHBY, O. L., and BOZORTH, R. M.: 'A New Magnetic Material of High Permeability', *Journal of Applied Physics*, 1947, **18**, p. 173.
(6) ASSMUS, F., and PFEIFER, F.: 'Höchstpermeable Legierungen im System NiFeCuMo', *Metall*, 1953, **7**, p. 189.
(7) WESTERN ELECTRIC CO.: British Patents Nos. 638210, 678339, 681466.
(8) SCHUMACHER, E. E.: 'Metallurgy behind the Decimal Point', *Transactions of the American Institute of Mining and Metallurgical Engineers*, 1950, **188**, p. 1104.
(9) KERSTEN, M.: 'Theoretical Remarks on the Influence of Slight Heterogeneous Impurities on the Initial Permeability of Nickel–Iron Alloys' (Soft Magnetic Materials for Telecommunications, edited by Richards, C. E., and Lynch, A. C., Pergamon Press, London, 1953), p. 1.
(10) PETERSON, E., and WRATHALL, L. R.: 'Eddy Currents in Composite Laminations', *Proceedings of the Institute of Radio Engineers*, 1936, **24**, p. 275.
(11) EPELBOIN, I., and MARAIS, A.: 'Texture magnétique macroscopique des rubans en Permalloy', *Comptes Rendus*, 1949, **229**, p. 1131.
(12) EPELBOIN, I.: British Patent No. 658425.
(13) 'Economy in Nickel–Iron Magnetic Alloys', *Metallurgia*, 1952, **46**, No. 273 (July), p. 34.
(14) FELDTKELLER, R.: 'Spulen und Übertrager' (Hirzel, Zürich, 1949), p. 82.
(15) EPELBOIN, I., and ABGRALL, C.: *Comptes Rendus*, 1952, **234**, p. 1265.
(16) RANDALL, W. F.: 'Nickel–Iron Alloys of High Permeability, with Special Reference to Mumetal', *Journal I.E.E.*, 1937, **80**, p. 651.
(17) WILDE, H.: 'Die Vorteile der Gegeninduktivitätsmessbrücke bei ferromagnetischen Messungen', *Archiv der Elektrischen Übertragung*, 1952, **6**, p. 354.
(18) LYNCH, A. C.: 'The Assessment of Inhomogeneity in Thin Strips of High-Permeability Alloys', Reference 9, p. 183.
(19) WALTERS, R. E. S.: Reference 9, p. 188.
(20) FELDTKELLER, R.: 'Spulen und Übertrager' (Hirzel, Zürich, 1949), p. 102.
(21) CHALMERS, B., KING, R., and SHUTTLEWORTH, R.: 'The Thermal Etching of Silver', *Proceedings of the Royal Society of London*, A, 1948, **193**, p. 465.
(22) ABGRALL, C., and EPELBOIN, I.: 'Sur l'étude en fonction de la temperature de rubans minces en mumétal soumis à des champs alternatifs faibles', *Comptes Rendus*, 1952, **234**, p. 1265.
(23) WILLIAMS, H. J., SHOCKLEY, W., and KITTEL, C.: 'Studies of the Propagation Velocity of a Ferromagnetic Domain Boundary', *Physical Review*, 1950, **80**, p. 1093.
(24) MENYUK, N., and GOODENOUGH, J. B.: 'Magnetic Materials for Digital-Computer Components. I. A Theory of Flux Reversal in Polycrystalline Ferromagnetics', *Journal of Applied Physics*, 1955, **26**, p. 8.
(25) LEE, E. W.: To be published.
(26) LYNCH, A. C.: *Proceedings I.E.E.*, 1954, **101**, Part II, p. 667.
(27) TOMBS, N. C., and WELCH, A. J. E.: 'Thermodynamic Properties of Silicon Monoxide', *Journal of the Iron and Steel Institute*, 1952, **172**, p. 69.

(9) APPENDICES

(9.1) Core Making

Although the primary purpose of this work was to study how to make a high-performance metal, it is clear that, if this performance cannot be realized in cores made from the metal, much of the work is wasted.

(9.1.1) Insulation.

Laminations of metal 50 microns thick and less cannot be conveniently handled, and the most practical alternative is the clock-spring core. With this type of construction the interlaminar insulant must be applied to the hard-rolled metal before it is wound into cores; it must therefore withstand firing in hydrogen at 1 000° C or over. Thin paper interleaving has been used commercially, the carbon residue after heat treatment being a good enough insulant for the purpose; unfortunately, however, the thinnest available paper is 8 microns thick, so that the heat-treated cores have a poor space factor, and since the paper shrinks when charred, they are always loose.

Other possible insulants are talc, silica, alumina, magnesia, chromic oxide, etc. Preliminary experiments with sprayed coatings on flat rings showed that all degraded the metal more or less and that the amount of degradation varied roughly as the

thickness of the insulant. This degradation is believed to be due in part to strains induced in the metal by the coating on cooling, since a good insulant adheres strongly to the metal but has a different coefficient of thermal expansion.

Aluminium oxide was one of the best, and coatings of it, if thin enough, caused no measurable falling-off in performance. These coatings could be applied either cataphoretically or by simple drag from a suspension of the oxide in a suitable organic fluid. Methyl alcohol containing a little nitrocellulose is a convenient vehicle, but many others can be used. Since interlaminar insulation does not need to be very good—a resistance of an ohm or so per turn is enough—very thin coats are adequate and a good stacking factor is possible.

Most commercial aluminas contain a little silica; this is an undesirable impurity for the reasons given in Section 4.1.2, and much better permeabilities are attained if the silica content is very low—say 10 parts per million.

Table 9 gives some typical results. The last two columns refer to the properties of the metal, not the completed core, i.e. the permeability must be multiplied by the space factor to give the core permeability.

Table 9

RELATION BETWEEN PERMEABILITIES OF STRIP AND OF CORES MADE FROM IT

Material	Thickness	Tested as rings		Tested as spiral core	
		μ	g	μ	g
	microns				
Commercial ..	50	39 000	1·33	34 000	1·67
Precision	30	60 000	2·2	42 000	2·2
	30	51 000	1·9	43 000	2·6
	30	44 000	1·5	35 000	3·3

The difference between the permeabilities of the precision materials is the result of deliberate differences in processing, particularly in the annealing temperatures; but the difference should in any case be regarded in the light of the final remark in Section 2. The high values of g show that there has been some short-circuiting of turns of the tape.

(9.1.2) Protection of the Core.

The final step in core making—and an essential one—is to protect the core against mechanical damage during winding. This is best done by enclosing it in a box, which may conveniently be a plastics moulding. The older practice, which was, however, intended mainly to fix the turns in position and prevent loss of insulant, was to solidify the coil by impregnating with wax or synthetic resin, and a modified form of this process may still be acceptable. The objection to the full impregnation process is that the contraction of the impregnant, not on solidification but on subsequent cooling, stresses the metal so much that the permeability is seriously reduced. Some experiments were carried out by making up some small cores of 25-micron materials and measuring their permeabilities whilst they were impregnated and cooled to room temperature. Typical results are given in Table 10, which shows trouble to some extent with every material which sets hard. Some data for thermal expansions are quoted, although the authorities do not agree about them.

The mere presence of an impregnant is not, however, harmful, as is shown by the result for petroleum jelly. A soft material such as polyisobutylene seems to be permissible, and a hard material does not do much damage if it is used to give a protective skin and is not allowed to penetrate the core fully. But either

Table 10

EFFECT OF VARIOUS IMPREGNANTS ON PERMEABILITY OF CORES

Impregnant	Temperature at which impregnant became solid	Coefficient of linear thermal expansion per °C	Loss of permeability on cooling to room temperature
	°C	$\times 10^{-6}$	%
Paraffin wax	54	110	40
Carnauba wax ..	84	320	80
Beeswax	62	250	50
Microcrystalline wax ..	74	Not known	50
Rosin + resin oil, 2 : 1 ratio			40
Polyisobutylene:			
M.W. 120 000, from solution at 60° C	60		30
M.W. 120 000, from solution at room temperature	Room temp.		2
M.W. 100 000, from solution at room temperature	Room temp.		9
Petroleum jelly ..	—		0

treatment is likely to give a core having a temperature coefficient of permeability which is large and dependent on the impregnant rather than the metal.

(9.2) The Chemical Significance of Silicon in Nickel-Iron Alloys

Traces of silicon may find their way into magnetic alloys from several sources. These small quantities, either in commercial melted or in sintered alloys, do not seem to affect the magnetic properties directly, though they may perhaps change the optimum cooling rate.[8] One significant effect, however, is to make the metal sensitive to the presence of water vapour in the annealing atmosphere, shown most strikingly in the way metals with and without silicon respond to small amounts of water vapour in the hydrogen during the final anneal. The following short discussion attempts to explain why traces of silicon can be so important.

When silicon and silicon dioxide (silica) are in contact with hydrogen and water vapour at high temperatures, these four reactions must be considered:

(i) $SiO_2 + H_2 \rightleftharpoons SiO + H_2O$.
(ii) $SiO + H_2 \rightleftharpoons Si + H_2O$.
(iii) $SiO_2 + Si \rightleftharpoons 2SiO$.
(iv) $H_2O \rightleftharpoons 2H + O$ (at the surface of the metal).

In all these reactions increasing temperature favours the reaction going from left to right. Increasing moisture in the annealing hydrogen favours the reverse reaction in (i) and (ii). It has no direct effect on (iii), but of course reduces the amount of free silicon available for the reaction. At high temperatures silica exposed to dry hydrogen is therefore reduced, first to silicon monoxide which is volatile, and then to metallic silicon.

The curve in Fig. 8,* based on data from Tombs and Welch,[27] shows that, at 1 350° C, hydrogen having a dew-point below −20° C tends to produce silicon monoxide from silica, whereas at 1 100° C only a much drier gas will permit this reaction. An alloy, originally free from silicon, is therefore likely to become contaminated by gaseous transfer if annealed in dry hydrogen at, say, 1 300° C in the presence of silica. With wet gas the pick-up should be much less. When the silicon monoxide reaches the metal surface it is either reduced by the metal itself or by hydrogen and is therefore deposited on the surface. Once there, it is rapidly dissolved by the metal and dispersed in it. The exact

* The extrapolation in this Figure is not due to Tombs and Welch and is not necessarily accurate.

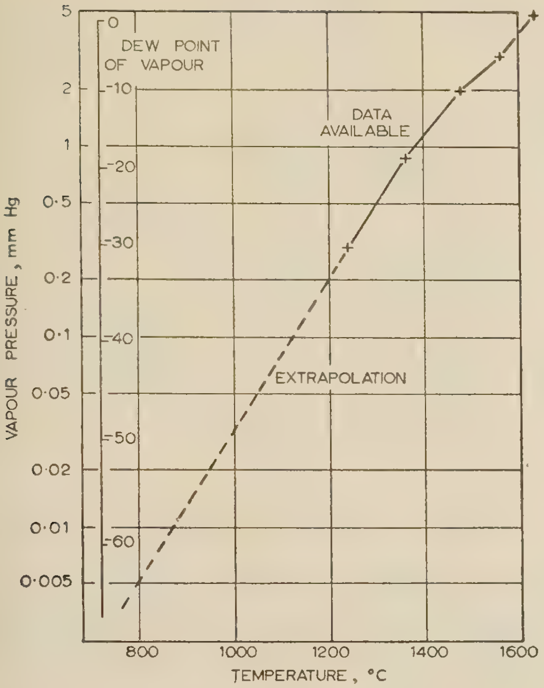


Fig. 8.—Water content of hydrogen in equilibrium with SiO_2 and SiO (after Tombs and Welch[27]).

mechanism of this is not important; what matters is that the pure metal has been converted into a dilute solution of silicon or a silicide.

Assume, for the moment, that silicon in dilute solution or combined as silicide will behave quantitatively as pure silicon and that the data in Fig. 8 are therefore relevant, and consider a silicon-free alloy being given a final anneal at 1 300° C in hydrogen of dewpoint −30° C. When, during the heating stage, the temperature passes 1 200° C any silica present will tend to be reduced to silicon monoxide, some of which will reach the metal and then be further reduced to silicon and dissolved. During the cooling after high-temperature treatment, when the furnace temperature has fallen below 1 200° C, the atmosphere, instead of reducing silica, will oxidize any silicon to silicon monoxide and silica. Hence any silicon or silicide in the metal, particularly in the surface layers, will be re-oxidized to silica either directly by water vapour [reverse of (i) and (ii)] or by atomic oxygen formed at the metal surface by thermal breakdown of water (iv). This silica, being insoluble in the metal, will be precipitated and set up internal strains, limiting the magnetic permeability.

This argument applies to alloys which were, at the beginning of the final anneal, free from silicon in any form. Practical alloys made by powder metallurgy, whether made in the laboratory or the factory, always contain traces of silicon and may be expected to react slightly to water vapour. Melted alloys are subject to the same influences but to an increased degree; they generally contain more silicon and have been melted under reducing conditions at temperatures substantially higher than those reached by sintered alloys. Most melted alloys also contain

Table 11

Joint Effect of Silicon Content and Annealing Treatment on Permeability

Material	Sintering temperature and state of gas	Annealing treatment: 3 hours at temperature	Annealing treatment: Humidity	Result: Silicon content	Result: Permeability	Additional annealing treatment: Temperature	Additional annealing treatment: Humidity	Result: Permeability
	°C	°C		%		°C		
Powder-metallurgy alloy, with 0·05% added SiO_2	1 050 dry	1 050	dry		24 000			
			wet		24 000			
	1 350 dry	1 050	dry		46 000			
			wet		37 000			
	1 350 wet	1 050	dry		45 000			
			wet		36 000			
Pure powder-metallurgy alloy	1 350 dry	1 300	dry	0·016	61 000			
					69 000 73 000	1 050	dry wet	66 000 58 000
		1 300	wet	0·009	67 000			
					65 000 66 000	1 050	dry wet	68 000 66 000
	1 350 dry	1 050	dry	0·012	41 000			
			wet	0·009	38 000			
Commercial melted alloy		1 300	dry		54 000			
		1 300	wet		17 000			
		1 050	dry		25 000			
			wet		13 000			

highly reactive deoxidants, such as magnesium, which can reduce silica to silicon. These alloys should react even more strongly to water vapour.

The case discussed above is that of a metal during its final anneal. The same chemical considerations hold during intermediate anneals whilst the metal is being cold-rolled. The formation of silica at this stage is less important, because, during the recrystallization which occurs when cold-rolled metal is annealed, insoluble impurities tend to be aggregated and swept to crystal boundaries where they are relatively harmless.

In so far as the above argument is quantitative, the dilute solution of silicon in the alloy has been treated as though it behaved like pure silicon. Actually it is unlikely that this will be so; the activity of silicon in solution will be less than unity, so a greater concentration of water will be needed to reverse (ii) than is indicated in Fig. 8.

This theory leads to the following predictions, in which 'wet gas' means gas represented by points above the curve of Fig. 8; at temperatures below about 1 000° C, this implies that any but the most rigorously dried gas is wet.

Low-temperature sintering.—Alloys with or without silica which have never been raised to temperatures above about 1 050° C should be insensitive to the presence of water vapour in the atmosphere during a final short anneal at 1 050° C. A long anneal in dry gas in the presence of silica will probably contaminate the metal with enough silicon to cause a fall in permeability.

High-temperature sintering (e.g. 1 350° C).

	Annealed in dry gas	Annealed in wet gas
Annealed at high temperature (e.g. 1 300° C)	Silicon if present will be picked up. Any silica in the alloy will be converted to silicon in solution or combination. The alloy will have a high permeability after this anneal but will be degraded by any subsequent annealing in wet gas	*Alloys free from silicon:* no effect; little or no silicon picked up *Alloys containing silicon:* degradation especially during cooling. Subsequent annealing at low temperatures in wet gas will have no effect
Annealed at low temperature (e.g. 1 050° C)	Silicon if present will be picked up slowly. If this continues too long, the alloy will be degraded by any subsequent annealing in wet gas	

Melted alloys.—These will behave in the same way as metals sintered at high temperatures, but owing to the higher temperature needed for melting, the effect will be more pronounced. The presence in commercial alloys of reactive deoxidants may further enhance this effect.

The experimental data in Table 11 tend to confirm the above predictions.

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL SECTION, 29TH JANUARY, 1957

Mr. W. F. Randall: The authors state that the properties of commercial alloys are inconsistent. Commercial nickel-iron alloys are, in their most generally used form, graded upwards from a specified minimum, known and used by designers, dictated by economics, and by the possible reuse of manufacturing scrap. This is a very big consideration when working to close tolerances. Equivalent standards of permeability have quite separately and independently been arrived at in Germany and the United States.

It is not made clear by the authors that this work has been carried out on very thin strip. The remarks and conclusions are not equally applicable to thicker gauges (up to 0·015 in) in which 90–95% of these alloys are used. The quantity of very thin strip used is a very small fraction of the whole, and in that range variations have been greater and permeabilities lower owing to the impact on their manufacture of factors not yet fully understood. The paper makes some contribution to that field.

The particular alloy used (77/14/5/4) was developed from the binary alloy, specifically with commercial and economic considerations in mind. Copper was added to spread the range of composition over which a reasonable permeability could be obtained and also to enable a simple heat treatment to be applied. These adjustments made the alloy an economic proposition. From the evidence in the paper, I find that established material compares well with the very elaborately prepared precision alloy.

The primary object of the work was economy in nickel, but it is very problematical whether this is achieved. As I have said, commercial production must allow reasonable margins, otherwise manufacturing scrap cannot be reused, and so, if we adhere to very close tolerances, the use of primary nickel will be increased rather than diminished.

The authors show clearly how small differences in strain, oxygen content, atmospheric conditions, temperature, rate of cooling, etc., can cause variations in the alloy. Only by the very elaborate precautions described in the paper has it been possible to maintain some consistency. In view of this, manufacturers should be congratulated on having maintained such a standard over so many years.

In Section 4.1.2 the reference to purity presumably applies to the precision alloy, and judging by Fig. 1 these very elaborate precautions are hardly having a profound effect, since the permeability developed in the two cases is much the same.

The penultimate paragraph of Section 4.1.2 contains a statement which I find difficult to reconcile with the evidence of Table 1. After 17 hours of annealing, which is surely a very long time, the silica tube gives a permeability of 50 000 and the alumina tube gives a permeability of 48 000. Admittedly, after 17 hours both fall away a little, but I should like further evidence before accepting the statement that the silica tube is responsible for this difference. In fact, in both Tables 1 and 4 there is an indication of deterioration after prolonged annealing.

Do the authors consider that the effect in Table 2 is due to the increase of diffusion rates and the promotion of greater homogeneity in the alloy?

In Fig. 6 the advantages of precision alloys for ultra-thin strip are demonstrated, which is their most obvious field of application. There is a maximum between 50 and 20 microns. Have the authors any theory to explain that maximum? Is it due at all to any rolling differences? In the same Figure there is a gradually declining permeability of the commercial alloy with diminishing thickness. The permeability at 200 microns is about 30 000. In Table 1 examples are cited with permeabilities of 39 000. Could a sample of that nature follow the same slope, because if so, it would fall above the lower curve in Fig. 6?

The permeability of strained material in a nickel-iron alloy is uniformly low, whether we start with a permeability of 70 000

or 20 000. The net result is that strain on the high-permeability alloy has a much greater overall effect proportionately than on the lower-permeability alloy. This may be of considerable importance when the materials are subjected to commercial handling.

The paper demonstrates that the very elaborate, expensive and costly preparation of sintered alloys shows very little advantage, except in the very thin strip field, over the alloys at present available.

Mr. C. Gordon Smith: The authors describe a rather new technique for the production of improved high-quality materials, which appears to offer considerable practical possibilities. It is unfortunate that the authors did not make it clearer that the use of these very-high-quality alloys is possibly somewhat restricted, since Mr. Randall's remarks might then have been less discouraging. Manufacture of high-quality materials is inevitably a costly process, but sometimes people may be prepared to pay for it.

Although Supermalloy originated in the United States ten years ago, availability, in this country at any rate, is very restricted. The position has improved somewhat recently with the appearance of Supermumetal in this country and Ultraperm on the Continent, but nevertheless there is still further scope for better materials.

With regard to the effect of composition, the authors have shown a peak in the nickel-iron ratio, keeping copper and molybdenum fixed. Can they give any further information on other compositional variations which might allow an assessment of the merits of the many variants of this type of alloy commercially available in different countries?

With reference to Section 5.3, the fact that the properties cannot always be restored by heat treatment is not generally known and ought perhaps to be explored further. Is it also to be assumed that, as the magnetic properties are further improved, their sensitivity to stress becomes greater? In many applications we can treat materials with due precautions, but in practice they are often liable to receive very rough treatment. Would it be possible to evolve another composition which, although showing moderately good magnetic properties, would be much less stress-sensitive?

Mr. S. E. Buckley: The two main types of impurity which affect the permeability of these alloys are sulphur, carbon, oxygen and nitrogen in one group, and elements such as aluminium, magnesium and silicon, which have a high affinity for oxygen, in the other. By the use of powder metallurgy these impurities can be greatly reduced and controlled both in the original materials and from subsequent pick-up from the processes. There is no melting, and so there is no pick-up from furnace linings or the atmosphere. Since all heating processes are carried out in hydrogen, some of the impurities are successively reduced, and no solid deoxidants are used, thus avoiding the introduction of magnesium or silicon.

The authors rightly point out that the advantages of high permeability must be considered in relation to cost. About 20 years ago, it was shown that iron could be produced with an initial permeability of 10 000, but there has been no commercial production because of the very high cost. The authors also point out that they do not expect high-quality alloys to replace all the present nickel-irons, but it is expected that they will be restricted to those applications where high permeability is especially required. High permeability can be obtained by a number of ways. It is often obtained now by selection from current production.

Another method follows the Supermalloy process. For Supermalloy, melting is carried out *in vacuo*, the ingots are rolled by normal commercial methods, and the final heat treatment consists of high-temperature purification in hydrogen followed by special heat treatment to produce a critical state of atomic ordering. The future will show which method of production is the most economic.

The point raised by Mr. Smith about the effect of strain on these alloys is important. Experience shows that the higher the permeability the greater the care that must be taken to avoid strain. It seems very doubtful whether the very-high-permeability alloys will ever be freely usable in the form of laminations. It is preferable to use spiral tape cores, and since the technique required is very specialized, they are better supplied as finished cores.

Dr. L. G. A. Sims: The paper is intended primarily for metallurgists rather than engineers. The qualities needed in magnetic materials can be summed up as high permeability, low loss, consistency of characteristics and non-ageing. The authors, however, do not seem to refer to the ageing characteristics, but since the work was spread over about ten years, I assume they have had experience of that particular property. Can they therefore provide some information on this matter?

The paper seems to imply that initial permeability is a criterion not only for consistency but also for performance in other fields. Is initial permeability a reliable criterion for the behaviour of the materials at finite exciting forces?

It is interesting to me to find the physicist's unit, the micron, now in common use for sheet magnetic materials.

Mr. G. Kerr: It may be of interest that there has been in existence for some time a production unit manufacturing this type of alloy by powder-metallurgy methods.

It has not been difficult to keep within the limits specified in the paper—oxygen, carbon and silicon to 0·01%, possibly 0·02% maximum, and for the iron we find that we can keep well within the limits. We feel confident, both by the strict control imposed on manufacture and the final results, that we are keeping to the requirements suggested as necessary by the authors.

Tables A and B show that we have obtained a fair degree of consistency in the material we produce.

Table A indicates the result for seven separate mixes of 110 lb each. Table B provides an examination of three in some detail, each compact being 11 lb. The material was tested in the form of rings at a frequency of 300 c/s, and the results have been adjusted to compensate for losses. Where two compacts bear the same number they refer to samples taken from opposite ends of the same compact.

Table A

Mix	Initial permeability of thickness of	
	0·005 in	0·004 in
A	35 000	
B	34 000	
C	38 000	
D	37 000	
E	35 000	38 000
F		33 000
G	30 000	

With reference to the controversy about hydrogen and cracked ammonia for annealing the specimens before testing, we have found no significant difference between results obtained on material annealed in these gases.

Dr. H. H. Scholefield: Fig. A shows the hysteresis loop for 15-micron Mumetal strip wound into a spiral core and heat treated. The d.c. initial permeability of this core measured at a

Table B

Compact No.	Mix E		Mix F	Mix G
	0·005 in	0·004 in	0·004 in	0·005 in
1A		35 000	28 000	34 000
1B		37 000		
2		45 000	33 000	34 000
3		35 000	39 000	29 000
4		39 000	35 000	33 000
5	33 000		33 000	30 000
6	32 000		36 000	30 000
7	37 000		33 000	30 000
8	32 000		31 000	32 000
9	42 000		33 000	30 000
10A	32 000		31 000	30 000
10B				33 000
Maximum	42 000	45 000	39 000	34 000
Minimum	32 000	35 000	28 000	29 000
Mean ..	35 000	38 000	33 000	30 000

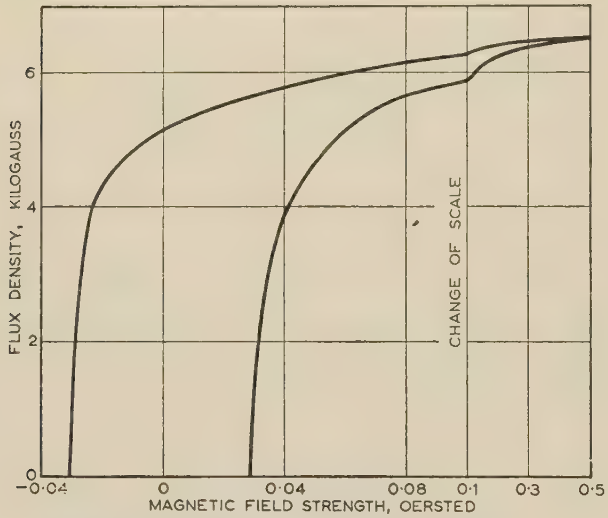


Fig. A.—Hysteresis loop for 0·000 6 in Mumetal.

Specimen: Spiral core 1½ in (outer diameter), 1 in (inner diameter).
Strip: 1¼ in wide, 0·000 6 in thick.

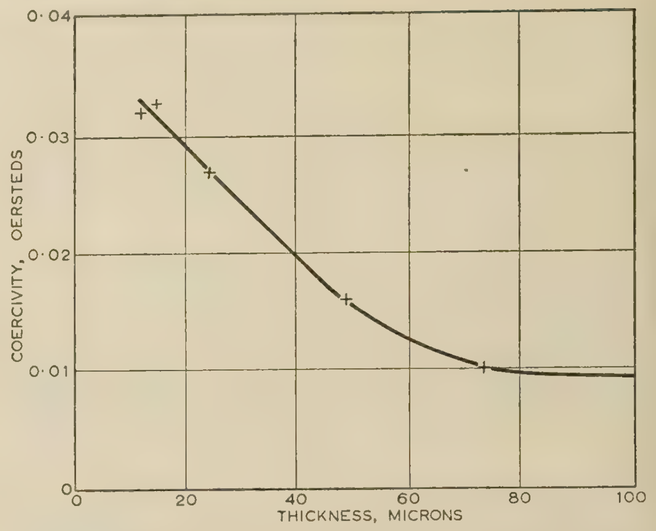


Fig. B.—Variation of coercivity with thickness for Mumetal.

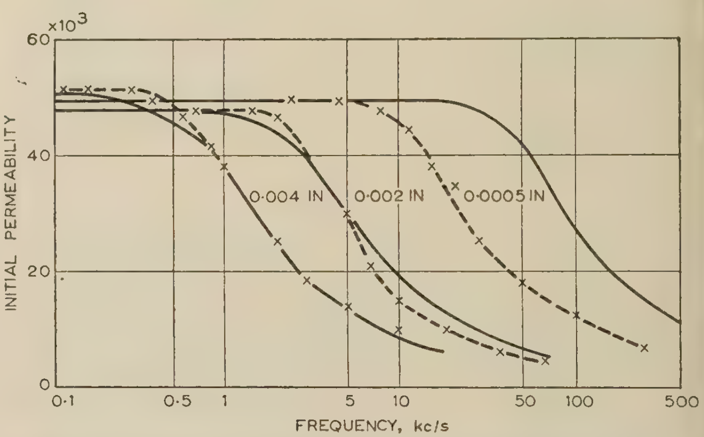


Fig. C.—Relationship between initial permeability and frequency for strip.

—— Theoretical.
---- Observed.

flux density of a few gauss was 38 000. The coercive force is 0·03 oersted, which is much higher than would be expected for material of this quality.

Fig. B shows the relationship between coercivity and strip thickness, and it will be noted that the coercivity increases from 0·01 oersted for strip 80 microns thick to 0·035 oersted for strip 10 microns thick. This curve refers to cores having an initial permeability in the region of 40 000. The increased losses which result are proportional to frequency, not the square of frequency, and as the authors state, they do not contribute much to the total loss at the frequencies at which the cores are likely to be used in many applications.

Fig. C shows the relationship between initial permeability and frequency for strip 0·000 5, 0·002 and 0·004 in thick. The dotted curves represent a plot of the experimental results obtained on a Maxwell bridge. The theoretical curves have been calculated from the formula due to Cauer which is as follows:

$$\mu_f = \frac{\mu_{f=0}}{\theta}\frac{\sinh\theta + \sin\theta}{\cosh\theta + \cos\theta};\ \theta = 2\pi t\left(\frac{\mu_{f=0} f}{\rho}\right)^{1/2}$$

The theoretical and practical results are in agreement for strip 0·002 and 0·004 in thick, whereas for a strip 0·000 5 in thick there is a marked disparity, which is associated with low a.c. resistivity, i.e. high value of g and anomalously high eddy-current losses. The phenomenon is referred to in Section 5.2 of the paper.

Fig. D illustrates the low anisotropy of present-day Mumetal. Two cores were prepared, one of strip with random orientation and the other of strip with preferred orientation (simple cube). The strip came from the same cast, and the final cores were heat treated together. The normal induction curves are almost identical, as indeed were the hysteresis loops.

Supermumetal cores are available with a minimum metal initial permeability of 50 000. These can be produced either from vacuum-melted or sintered material.

The cost of producing strip from sintered blocks is considerable, and attention has been directed to continuous compacting techniques followed by continuous sintering of the 'green' compact to give it adequate strength for handling and the subsequent high temperature heat-treatment/rolling schedule.

However, approximately two-thirds of the material produced by melting conforms to the composition requirements for optimum magnetic performance, so that specifications of minimum

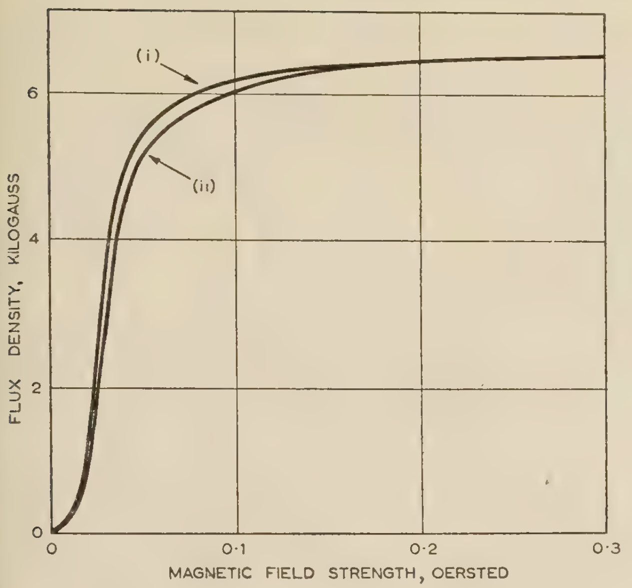


Fig. D.—Effect of preferred orientation on the normal induction curve of 0·0005 in Mumetal.

(i) Preferred orientation absent.
(ii) Preferred orientation present. Flux parallel to rolling direction [100].

magnetic quality can be fairly readily met by selection in those instances where this is a significant advantage to the designer.

Dr. G. A. V. Sowter: It is stated that 'designers of electrical equipment normally assume properties greatly inferior to those usually quoted'. So far as the manufacturers of Mumetal are concerned, materials are sold against purchasing specifications with minimum guarantees, and no assumptions of greatly inferior properties are made by designers utilizing this alloy.

The accusation might be construed to refer to commercial alloys of any thickness and in any form; the paper is devoted to thin strip only, of the order of 0·002 in or less. The whole of the current-transformer industry in this country has been operating on minimum-guarantee performance over a range of induction for many years, and I have yet to meet an instance where a customer is of the same opinion as the authors.

It is also stated in the Summary that 'spiral cores, if free from strain, can have properties almost as good as those of the strip they are made from'. In Table 9, in which ring tests are considered to represent strip, relative properties of precision rings and spiral cores are, in one instance, 60 000–42 000 and in another 44 000–35 000, which can hardly be regarded as equivalent.

The authors and a speaker finally consider 30 000 as the commercial metal permeability for 0·002 in strip, which, if corrected for the true difference between 'metal' and spiral cores, is not greatly dissimilar to the 20 000 for spiral cores offered commercially, as indicated in Section 1.

I am still puzzled by the remark that 'a core known to have a permeability of 15 000 can be smaller than one which might have a permeability of only 5 000, *a very usual design figure*'. Yet in Section 1 it is stated that two manufacturers during 1955 began to offer cores having permeabilities of 20 000 or more. It is time that the designers who operate on a figure of 5 000 took advantage of the manufacturers' guarantees. Furthermore, these designers should be aware of Supermumetal, which is available in the form of spiral cores.

Mr. J. J. Hill (*communicated*): The introduction of nickel-iron alloy in place of silicon-iron revolutionized the manufacture of precision current transformers. At the N.P.L. there is a range of first-class 50 c/s current transformers made of selected commercial cores of this material. However, we welcome the introduction of alloys with consistently higher permeabilities, because it will enable us to improve the performance of transformers, and by reducing the ampere-turns required, to increase the frequency range.

A current transformer has recently been constructed of ratio 0·25/5 amp, using a spiral core of initial permeability 45 000. Owing to the small number of turns used, the dielectric admittances are unusually low, and the transformer is regularly used for precision measurements over the frequency range 50–500 c/s. Experimental work is being carried out on low-level voltage transformers wound on spiral cores having permeabilities in the range 20 000–35 000 for strip thicknesses of 25–125 microns. These transformers will be operated over the range 50–10 000 c/s.

The production of strip with high permeability and of thickness down to 6 microns will be of great use at the higher frequencies, where eddy-current losses may be important.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. C. E. Richards, E. V. Walker and **A. C. Lynch** (*in reply*): Nearly all the speakers are producers rather than users of nickel-iron alloys, and some of them may not have understood the purpose of the paper. We make no criticism of the general level of the quality of high-permeability nickel-iron alloys; we lament only the occasional batch, inferior either as it comes from the original manufacturer, or as it leaves the manufacturer of laminations or spiral cores, which forces the users to adopt their very low design values. We decided to study, first, the material at the stage at which it is supplied for making into laminations or spiral cores, and immediately found wide variations in properties; for example, the 33 commercial alloys obtained in or about 1951, for which Fig. 1 gives data, were not specially chosen in any way, and while the general average was satisfactory, they included eight having permeabilities of under 20 000. (These 33 alloys were from three different manufacturers.) This called for further investigation, which we have tried to carry out. Our conclusions are that such variations could be largely avoided, but, as we point out at the end of the paper, this may not be economically justifiable for the thicker materials. Thin strip, which is expensive anyway because of the high cost of rolling, ought to be made only from material of high quality, but this condition has not always been met.

In reply to Mr. Randall, we think that Table 2 and Fig. 6 both show the effect of purification at high temperatures. Melting may be even more effective than this heat treatment for removing some of the impurities (although unfortunately it is likely to introduce others), and that is why, at the greater thicknesses, the melted alloys show the higher permeabilities.

Some of our recent work suggests that, among batches of a given type of alloy, the initial permeability is a good guide to the high-field properties. We have no information of our own on ageing, but Mr. Hill states that one of his cores showed no measurable change of properties over 12 months.

We have not explored the whole range of compositions about which Mr. Gordon Smith asks, but we are inclined to agree with Mr. Randall that the tolerance on composition is improved by the presence of copper—whether in the small proportion which Mr. Randall introduced so many years ago in Mumetal or in the much larger proportion used in 1040 Alloy.

We are grateful to Dr. Scholefield for his confirmation of the high a.c. loss which we had found in thin strip and of the absence

of directional effects in modern Mumetal. The high coercivity of thin strip, shown in Fig. B, agrees with that reported by Russian workers. The formula which he quotes from Cauer is for the series permeability; we have preferred to use the parallel value, which varies less with frequency.

We agree with Mr. Buckley's remarks. The effect of strain is undoubtedly the main disadvantage of this type of alloy; everyone would welcome the material whose properties Mr. Gordon Smith outlines, but we cannot suggest how to make it.

Both Mr. Randall and Dr. Sowter imply that there are agreed standards of performance to which commercial alloys conform, and yet there is no British Standard governing their magnetic properties; the draft referred to by Halsey[3] in 1949 still remains a draft eight years later. Admittedly the position is improving; the manufacturers who agreed to guarantee a permeability of 7500 for a core of 50-micron strip in 1952 would probably guarantee a higher value now. But the much higher values mentioned in the discussion have been made possible only by the adoption of powder metallurgy as a manufacturing process or by other techniques so special that the resulting material is sold under a name different from that of the normal alloy. It would be helpful to users if Dr. Sowter would publish the minimum values to which his ordinary product conforms; at present, the only data generally available are 'typical' values which are useless for design purposes.

We are gratified by the results reported by Mr. Hill for cores made from powder-metallurgy material, and we hope to develop material of yet higher permeability for this application.

621.317.335.3

Paper No. 2208 M
Oct. 1956

A METHOD FOR THE PRECISE MEASUREMENT OF PERMITTIVITY OF SHEET SPECIMENS

By A. C. LYNCH, M.A., B.Sc., Associate Member.

(*The paper was first received* 12*th May, and in revised form* 9*th August*, 1956. *It was published in October*, 1956, *and was read before the* MEASUREMENT AND CONTROL SECTION 29*th January*, 1957.)

SUMMARY

A specimen, not itself carrying electrodes, is inserted loosely in a prefabricated electrode assembly and the capacitance is measured. The specimen is then removed and the electrode separation reduced until the system has the same capacitance as before. The permittivity can then be calculated without knowledge of the electrode area or of the calibration of any capacitors. The precision of the method is limited only by the difficulty of measuring the thickness of the specimen, and errors may be as low as 1 in 1 000. For measurement of power factor this method is less sensitive than others, but it is free from uncertainty about effects introduced by electrodes. It can be used at frequencies up to about 1 Mc/s, and being quick in use is suitable for routine measurements of a batch of nearly similar specimens.

(1) CURRENT PRACTICE

A recent paper by Rushton and Parry[1] sets out the normal techniques for measuring the permittivity and power factor of dielectric materials available in sheet form. It recommends using a disc of about 5 cm diameter and a few millimetres thick; the authors probably intend not more than 2 mm thickness, so that the capacitance obtained is of the order of 50 $\mu\mu$F. Such a capacitance can easily be measured to 1 part in 1 000, or better if necessary; the area of the electrodes can be defined within perhaps 1 part in 500; but the thickness of such a disc is difficult to measure to better than 1 part in 200, and this sets the limit to the accuracy of measurement of permittivity if the normal forms of apparatus are used. If the material is not quite rigid (e.g. polythene) even this accuracy is difficult to attain.

The same paper lists many types of electrode which have been used, finally recommending mercury. Other workers have favoured tinfoil or films of silver or aluminium deposited by evaporation. All these are, in their respective ways, troublesome to apply, and Rushton and Parry regard all of them as open to occasional suspicion. They do not mention the possibility of using electrodes separated from the specimen by an appreciable gap.

McCall[2] has very recently developed a method of measuring the permittivity of a sheet specimen by immersing it in a standard liquid of similar permittivity. His method has most of the advantages claimed for the one now to be described, but only if a suitable liquid exists.

(2) AN ALTERNATIVE METHOD

The present paper describes a method intended to allow measurement of permittivity with high precision; errors of only 1 part in 1 000 might be possible with suitable specimens. Although this precision would rarely be needed, it might be useful for the control of the composition of a mixture. The method avoids the need for applying any electrodes whatever to the specimen, and may be very desirable for this reason alone. Further, no pressure need be applied to the specimen, so that the method is suitable for soft materials.

Mr. Lynch is at the Post Office Research Station.

The main feature of the method is the use of a specimen so thick (e.g. 0·5 cm or more) that its thickness can be measured with the required accuracy, and the measurement or comparison of capacitances of the order of 10 $\mu\mu$F with an accuracy of about $\pm 0{\cdot}005\,\mu\mu$F, which is not difficult. The electrode system is prefabricated, and the specimen is merely inserted into the space between the electrodes (it may rest on one of them), the capacitances of the system being measured with and without the specimen in position.

Although measurements of this kind are in principle sufficient, the precision of the method can be greatly improved, and the calculations simplified, by arranging for the separation of the electrodes to be controlled by one or more micrometer screws; in this case the procedure is first to measure the capacitance, in some arbitrary units, when the specimen is between the electrodes, and then to remove the specimen and find how far to bring the electrodes together to produce the same capacitance as before. In practice the correct position for the second measurement is found by setting the plates to a few known positions, measuring the capacitance for each position (still in arbitrary units), and interpolating between the capacitances to find the exact position which would be required.

(3) CALCULATION

(3.1) Calculation of Permittivity

Let C_1, C_2 = Capacitances respectively with and without the specimen between the electrodes.

A = Area of the electrodes.

t_s = Thickness of the specimen.

t_o = Gap remaining between the specimen and the electrodes.

t_x = Reduction in separation of the electrodes between the two measurements.

ϵ = Permittivity of the specimen.

Then, in C.G.S. units,

$$\frac{1}{C_1} = \frac{4\pi}{A}\left(\frac{t_s}{\epsilon} + t_o\right)$$

and

$$\frac{1}{C_2} = \frac{4\pi}{A}(t_s - t_x + t_o)$$

When t_x has been adjusted so that $C_2 = C_1$,

$$\frac{t_s}{\epsilon} = t_s - t_x$$

or

$$\epsilon = \frac{t_s}{t_s - t_x} \qquad (1)$$

This expression does not involve either areas or capacitances.

(3.2) Calculation of Power Factor

The method was not primarily intended for measuring power factor, but nevertheless may be so applied. Its sensitivity is less than that of other methods because the power factor is 'diluted'

by the presence of the air-gap, but, on the other hand, there can be no error due to poor contact between electrodes and specimen.

Let C_s = Capacitance of the part of the specimen which is between the electrodes.
R = Equivalent series resistance associated with it.
ΔG = Measured decrease in conductance between the two balance conditions.
t_a = Final separation of the electrodes (so that $t_a = t_s - t_x + t_o$).
f = Frequency.
$\omega = 2\pi f$.

Then the power factor of the specimen is

$$\omega C_s R = \omega C_1 R \frac{C_s}{C_1}$$

$$= \frac{\Delta G}{\omega C_1} \frac{C_s}{C_1}$$

provided that the power factor is small, so that the equivalent series and shunt capacitances can be assumed equal. Since $C_2 = C_1$, the power factor is

$$\frac{\Delta G}{\omega C_2} \frac{C_s}{C_2}$$

$$= \frac{\Delta G}{\omega C_2} \epsilon \frac{t_a}{t_s}$$

$$= \frac{\Delta G}{\omega C_2} \frac{t_s}{t_s - t_x} \frac{t_a}{t_s}$$

$$= \frac{2\Delta G}{A} \frac{}{f} \frac{t_a^2}{t_s - t_x} \times 9 \times 10^{11}$$

if ΔG is in mhos and the lengths are in centimetres.

(4) APPARATUS

(4.1) Electrode Assembly

Since the capacitances to be measured are necessarily small, and stray capacitances are therefore comparable with them, one of the electrodes is fully screened; and, for reasons given below, the working area is closely defined by the use of a guard ring. Fig. 1, a section through the electrode assembly, shows how this can easily be arranged. The high-potential electrode is mounted on the base-plate by three short insulating pillars; the guard electrode stands on the base-plate on three micrometer spindles provided with steel cups and balls. Two of the balls stand in a hole and a slot, so that the electrode system is located by the 'kinematic design' method. The guarded electrode is mounted from the guard electrode, and therefore has nothing except air between it and the high-potential electrode. The guarded and guard electrodes should be machined after assembly, so that their surfaces are in the same plane. The micrometer-head mountings should be so adjusted that, for equal readings on their scales, the electrodes are nearly parallel. Short studs are fitted on the lower electrode, well outside the working area, to locate the edge of the specimen.

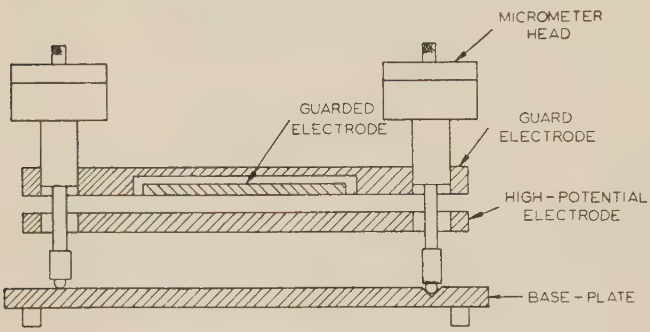


Fig. 1.—Sectional view of essential parts of apparatus.
For clarity, supports for the electrodes are omitted from the diagram.

The apparatus used for the experiments described in the paper has a guarded electrode of 3 in diameter, in a recess $3\frac{1}{16}$ in in diameter. The other electrodes are 7 in squares. An 8 in square would have been a better choice because it would have allowed a wider range of sizes of specimen to pass between the micrometer spindles. The micrometer heads have barrels 2 in in diameter, graduated to 0·000 1 in. They indicate approximately 1·000 0 in when the plates are in contact. The apparatus needs a specimen either about $4\frac{1}{4}$ in in diameter or about $4\frac{1}{2}$ in wide and of any length greater than $4\frac{1}{2}$ in.

For measurements over a very wide range of frequencies, the specimen might be made interchangeable with that needed for the reactance-variation apparatus developed by Hartshorn and Ward.[4] The diameter of the specimen should then be $2\frac{3}{8}$ in and that of the working electrode a little less than 2 in. The specimen would probably have to be thinner than those recommended here, but its permittivity would still be determined with an accuracy limited only by the measurement of its thickness. To minimize the error discussed in Section 6.1, the apparatus intended for a specimen so little larger than the guarded electrode should be constructed with the high-potential electrode above the others.

(5) MEASUREMENT OF THICKNESS OF SPECIMEN

For the desired overall accuracy of 1 part in 1 000, the thickness of the specimen must also be measured with this accuracy. It is difficulty to find anything better than an ordinary micrometer for this purpose. Even on a yielding material such as polythene, any one person can, with a little practice, obtain results reproducible within about 0·0002 in, and the variation between the results of several observers will not be much greater if they all use similar techniques. A good plan is to use a jig to ensure that measurements are made at points suitably distributed over the area to be tested. If an ordinary micrometer is used on a yielding material, the ends of the anvils should be of $\frac{1}{4}$ in diameter, so that the pressure applied does not become excessive. A micrometer designed at the N.P.L. for work of this kind, however, has a small non-rotating anvil and means of controlling the contact pressure. Pneumatic gauging has also been tried, and is promising.

(6) EFFECTS OF IMPERFECTIONS IN APPARATUS AND SPECIMEN

It has been assumed so far that the field is everywhere perpendicular to the electrodes, and that all planes parallel to the electrodes are equipotentials. This may not be realized in practice.

(6.1) Distorted Field at Edge of Specimen

Ideally, the specimen would be very large, so that any discontinuity in conditions at its edge could not affect the measured capacitance. The error introduced by the finite size of specimen is difficult to calculate. Experiment shows that the error is least when the thickness of the specimen is almost equal to the separation between the electrodes and when the specimen touches the guarded rather than the high-potential electrode. The second of these conditions reduces the error by a factor of up

to 4 and reverses its sign, but in the apparatus described it was not complied with for fear of filling the gap between the guard and guarded electrodes with fragments of material from the edges of specimens. The error with a circular specimen resting on the high-potential electrode, overlapping the working electrode by an amount equal to its thickness, and filling about seven-eighths of the space between the electrodes, is of the order of 0·01 $\mu\mu$F, which gives an error of the order of 1 in 1 000 in the permittivity. The dimensions suggested in this paper give an overlap, however, of several times the specimen thickness. The use of a square or rectangular specimen further reduces the error.

(6.2) Gap between Guarded and Guard Electrodes

The effect of the annular gap between the guarded and guard electrodes was checked by using modified apparatus in which this gap had been increased to 0·094 in—i.e. three times its normal value. The permittivity of a specimen then appeared to be lower by one part in 1 500. With electrodes of the sizes described in Section 4.1, therefore, the error due to the annular gap is unlikely to exceed one part in 3 000.

(6.3) Electrodes not Parallel

The effect of the electrodes not being parallel was easily checked experimentally. The three micrometers were each set to 0·650 in (i.e. an electrode separation of about 0·350 in) and the capacitance of the system was measured; then two of the micrometers were set respectively to 0·675 and 0·625 in (so that the mean spacing was the same as before) and the increase of capacitance was measured. This increase was 0·005 $\mu\mu$F. As this increase is proportional to the square of the error in parallelism, the uncertainty in capacitance for any likely error in the initial adjustment of the electrodes (say 0·003 in) is of the order of 0·000 1 $\mu\mu$F, which is negligible.

(6.4) Surfaces of Specimen not Parallel

A non-uniform specimen introduces three sources of error. One effect is similar to that just described, though not so easy to measure experimentally. To obtain some idea of the error it can introduce, consider a specimen of which the thickness of one half is t_1, that of the other half t_2, and the mean thickness of the whole is t_s. Ignore the distortion of the field near the step in thickness. Eqn. (1) will then give a value which is too high by the factor $[1 + \frac{1}{8}(t_1 - t_2)^2/t_s^2]$. As an example, consider a specimen 0·25 in thick, with a variation of $\pm$0·005 in in various places in an area 3 in in diameter; the error in ϵ given by this expression is 1 part in 625, and the actual error will be less because the thickness will not everywhere deviate from the mean by the full 0·005 in.

The other two sources of error arise in the measurement of thickness. The part of the specimen whose mean thickness is measured may not be the same as the part under the working electrode; this is dealt with by defining the working area with a guard ring, as described above, using the stops to locate the specimen in a known position between the electrodes, and using a jig during the thickness measurement to define the same

Table 1

Measurements on Specimens of Polythene containing Butyl Rubber and Antioxidant

Frequency	Specimen	Micrometer settings	Bridge readings		Micrometer setting for same capacitance without specimen	Permittivity	Power factor
			C	G			
kc/s		in × 10^{-4}	arb. units	micromhos	in × 10^{-4}		
0·12	1	7 000	2 207	0·000 071 0	8 323	2·294	0·000 5
	2	7 000	2 108	0·000 071 0	8 307	2·294	
	—	8 300	2 068	0·000 073 0			
	—	8 310	2 125	0·000 073 0			
	—	8 320	2 185	0·000 073 0			
	—	8 330	2 244	0·000 072 9			
1	1	7 000	2 218	0·001 385	8 323	2·294	0·000
	2	7 000	2 123	0·001 386	8 307	2·294	
	—	8 300	2 082	0·001 389			
	—	8 310	2 140	0·001 389			
	—	8 320	2 201	0·001 389			
	—	8 330	2 259	0·001 388			
10	1	7 000	1 934	0·034 34	8 322	2·292	0·000 1
	2	7 000	1 843	0·034 34	8 307	2·294	
	—	8 300	1 804	0·034 38			
	—	8 310	1 863	0·034 38			
	—	8 320	1 923	0·034 38			
	—	8 330	1 983	0·034 38			
100	1	7 000	1 858	0·411 1	$8\,322_5$	2·293	$0{\cdot}000\,1_5$
	2	7 000	1 766	0·411 1	$8\,306_5$	2·293	
	—	8 300	1 726	0·411 5			
	—	8 310	1 784	0·411 6			
	—	8 320	1 843	0·411 6			
	—	8 330	1 904	0·411 5			
1 000	1	7 000	2 079	1·541	$8\,321_5$	2·291	0·000 1
	2	7 000	1 991	1·540	$8\,306_5$	2·293	
	—	8 300	1 952	1·547			
	—	8 310	2 011	1·548			
	—	8 320	2 070	1·548			
	—	8 330	2 129	1·547			

area as is used for the electrical measurement. The other error arises from the finite size of the micrometer anvil; if the two surfaces of the specimen are not parallel, the micrometer reading will be larger than it should be. For a ¼ in anvil and a 3 in-diameter working area, the error is of the order of one-sixth of the difference between the largest and smallest thickness found in this area. This can be used as a correction to the measured thickness, but if it is more than about 0·001 in, the specimen should be rejected.

(6.5) Conductivity of Surfaces of Specimen

If the specimen were truly flat and of uniform thickness, surface conductivity would do no harm, but in practice this cannot be assumed. A specimen thought to have conducting surfaces should not be allowed to touch the electrodes, but should rest on a sheet of insulating material (not necessarily of low loss), which can be left in place when the specimen is removed, and treated as a part of the electrode system.

(7) EXAMPLES

The method shows to greatest advantage when a number of similar specimens have to be measured, because the same group of measurements with specimen removed will then serve for all of them. To illustrate the precision obtainable with the method, however, it will be more interesting to quote measurements made on the same specimen at a number of frequencies. The observations summarized in Table 1 were made with two sheets of a polythene mixture, of thickness 0·2345 and 0·2317 in respectively. The bridge network was that described in the companion paper.[3] Both capacitance and conductance are measured from arbitrary zeros (the capacitance on a reversed scale); the capacitance is in arbitrary units of approximately 0·001 $\mu\mu$F and the conductance is in micromhos. The micrometer readings are in ten-thousandths of an inch.

The close agreement between the measurements at the various frequencies suggests that they are free from frequency-dependent errors.

It is clear from Table 1 that more specimens might have been measured without the need for any further measurements in the 'specimen-out' condition.

Table 2 compares results obtained by the technique here described and by a method based on immersion in methylcyclohexane. As the permittivity of the liquid itself had to be measured there may be accumulated errors in this method, and the results obtained by the movable-electrode method are probably the better. There is further evidence here that the permittivities of two specimens may not be the same.

Table 2

COMPARISON OF RESULTS FROM TWO METHODS

	Permittivity	
	By movable-electrode method	By immersion
Specimen 3 ..	2·282	2·288
Specimen 4 ..	2·290	2·293

(8) CONCLUSIONS

The method described in the paper has the following advantages over conventional methods for the measurement of permittivity.

(*a*) It does not require knowledge of the calibration of any electrical components.

(*b*) It does not require measurements of electrode area.

(*c*) There is no uncertainty about whether the electrodes are introducing some spurious effect.

(*d*) It is quicker, particularly if a number of similar specimens are to be measured.

It can be used at frequencies from about 100 c/s to about 1 Mc/s. At higher frequencies the stray impedances, inevitable in apparatus 7 in wide, cause errors. To some extent they can be allowed for, but the method then loses its simplicity.

The limit to the precision attainable is set entirely by the difficulty of measuring the specimen thickness, which should be as large as is consistent with flatness and uniformity. For permittivities of about 2 the precision is of the order represented by an error of 1 in 1 000; for higher permittivities the method is less precise. The sensitivity of the available electrical apparatus would allow measurement to 1 part in 10 000 if the thickness could be measured precisely enough.

For the measurement of power factor the method has the disadvantage of 'diluting' the quantity to be measured, because there is the capacitance of an air-gap in series with that of the specimen, but this is usually outweighed by the absence of any uncertainty about the effects of electrodes applied to the specimen.

As is shown in Section 6.3, no extraordinary precision is needed in constructing the apparatus.

(9) ACKNOWLEDGMENTS

In the author's original apparatus both electrodes were fixed, and metal plates were inserted to vary the capacitance. The suggestion that one electrode should be movable was made by Mr. P. L. Parsons.

Acknowledgment is made to the Engineer-in-Chief of the Post Office and to the Controller of H.M. Stationery Office for permission to publish this paper.

(10) REFERENCES

(1) RUSHTON, E., and PARRY, J. V. L.: 'The Measurement of the Permittivity and Power Factor of Ordinary Solid and Liquid Dielectrics', N.P.L. Symposium on Precision Electrical Measurements, November, 1954 (H.M.S.O., 1955).
(2) MCCALL, D. W.: Private communication.
(3) LYNCH, A. C.: 'A Bridge Network for the Precise Measurement of Direct Capacitance' (see next page).
(4) HARTSHORN, L., and WARD, W. H.: 'The Measurement of the Permittivity and Power Factor of Dielectrics at Frequencies from 10^4 to 10^8 cycles per second', *Journal I.E.E.*, 1936, **79**, p. 597.

[The discussion on the above paper will be found on page 366.]

621.317.335.2

Paper No. 2209 M
Oct. 1956

A BRIDGE NETWORK FOR THE PRECISE MEASUREMENT OF DIRECT CAPACITANCE

By A. C. LYNCH, M.A., B.Sc., Associate Member.

(*The paper was first received* 12*th May, and in revised form* 9*th August*, 1956. *It was published in October*, 1956, *and was read before the* MEASUREMENT AND CONTROL SECTION 29*th January*, 1957.)

SUMMARY

For measurements by substitution only, the balance of a transformer bridge network can be made independent of capacitance and conductance from either terminal of the specimen to earth. In the network proposed in the paper both oscillator and detector have one terminal connected to earth. The network is particularly useful for measuring capacitances of less than 100$\mu\mu$F in the presence of stray capacitances to earth. It can be used at frequencies between about 4 c/s and 20 Mc/s; at frequencies up to about 1 Mc/s it is suitable for work of the highest precision.

LIST OF SYMBOLS

A = A negative capacitance.
C = Capacitance to be measured.
C_m = Capacitance between windings of the transformer.
C_s = Self-capacitance of a transformer winding.
C_v = Capacitance of a variable capacitor.
C_1 = A capacitance to earth.
C_k = An auxiliary capacitance.
L = Leakage inductance of a transformer winding.
L_1 = Inductance of a connecting lead.
R = Resistance of a transformer winding.
R_v = Variable resistance } in a T-network.
R_1, R_2 = Fixed resistances } in a T-network.
ω = Angular frequency = $2\pi f$.

(1) THE REQUIREMENT

A capacitance which is to be measured precisely is usually that of a 'three-terminal' capacitor, i.e. the capacitor has a screen, which is normally connected to earth, and the capacitance to be measured is that directly between the two other terminals, ignoring the capacitances between each of these terminals and the screen.

For precise measurement of capacitance a substitution method is nearly always used. This may take either of two forms. In one, the unknown capacitance is replaced by a fixed capacitance nearly equal to the unknown, the small difference between the two being measured on a standard variable capacitor. In the other, the unknown capacitance is removed from the circuit and replaced by an increase in the capacitance of a standard variable capacitor. In either case, when the substitution is made the unwanted capacitances from various points to earth are changed, and the circuit used for measurement must be unaffected by these changes. This condition is particularly difficult to meet if the capacitance to be measured is small compared with the capacitances to earth, which are usually of the order of 1–100 $\mu\mu$F. This difficulty arose, for example, in the development of a method for measuring permittivity.[1] It is the same difficulty that is normally met by the use of a Wagner earth connection, but the paper shows how to overcome it with fewer components than for the Wagner earth and no more adjustments than those for the simple bridge network.

Mr. Lynch is at the Post Office Research Station.

Similar difficulties may arise in the measurement of resistance in the presence of capacitance, and a similar method can be used to overcome them.

For simplicity, the following description applies to capacitance measurement only, and in Sections 2 and 3 the question of phase (or power-factor) balance is ignored.

(2) CIRCUITS FOR MEASUREMENTS BY SUBSTITUTION

When a bridge network is used for measurements by substitution, several features which are otherwise important cease to be so. The oscillator, the detector and three arms of the network are used merely to show that a certain arbitrary condition of the network has been restored after the substitution. The solution of the bridge equations is not needed, and the components in the three arms mentioned need not be precisely adjusted to nominal values or free from 'residuals', though they must be stable. The substitution bridge is thus so different from the ordinary one that it may be better to design it anew rather than to adapt existing networks.

The essentials of a circuit for substitution measurements are shown in Fig. 1; C is the capacitance to be measured, C_v can be

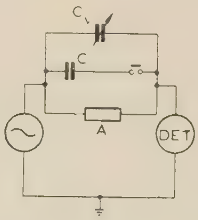


Fig. 1.—Essentials for a substitution bridge network.

adjusted to two values differing by C, and A, derived from an auxiliary 3-terminal network, is a negative capacitance which, in parallel with C_v, gives zero total capacitance so that no current flows from the oscillator through the detector. In this circuit, capacitances to earth fall across either the oscillator or the detector, and do not affect the measurement. The problem is how to realize the negative capacitance A. One solution is to use a twin-T network, but this makes the balance of the whole circuit strongly dependent on frequency, and both the oscillator and the detector must then meet special requirements.

The method now proposed is that the negative capacitance A should be obtained from a transformer and a capacitor, as shown in Fig. 2. The transformer is conveniently of 1 : 1 ratio (although this is not essential) and if so the associated capacitance C_k will be equal to the maximum valve of C_v.

The oscillator and the detector can be interchanged if desired. If this is done, however, an astatically-wound transformer should be used, to avoid errors due to stray magnetic fields. In the circuit shown in Fig. 3, any such fields generate signals which are substantially balanced in the network and do not reach the

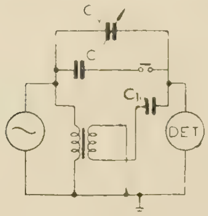


Fig. 2.—The proposed network.

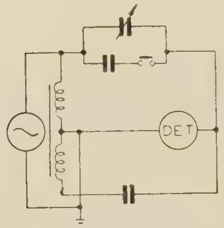


Fig. 3.—The same circuit as in Fig. 2, redrawn for comparison with Fig. 4.

detector, but in the conjugate network they flow directly through the detector circuit.

If the circuit is being used only to indicate equality of two capacitances, as in the application for which it was first designed, it is possible to replace the variable capacitor C_v by a T-network of three capacitors. The two which are not earthed can be of fixed values, while a variable capacitor is used between their junction and earth. Such an arrangement allows an ordinary two-terminal capacitor of about 1 000 $\mu\mu$F (maximum) to introduce changes as small as 0·001 $\mu\mu$F in the effective capacitance across the specimen. This device cannot be used in ordinary bridge circuits, because it also introduces large changes in other capacitances in the network; in the circuit described, however, these changes appear only across the source and the detector.

(3) COMPARISON WITH INDUCTIVELY COUPLED BRIDGES

The proposed circuit, which is redrawn in Fig. 3, is very similar to known networks[2, 3] described either as transformer bridges or inductively coupled ratio-arm bridges. With oscillator and detector interchanged, it has some resemblance to a network invented much earlier.[4] Fig. 4 shows one such circuit; in other

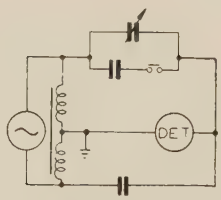


Fig. 4.—A bridge network with inductively coupled ratio-arms.

circuits the oscillator is connected to an additional winding on the transformer. There is, however, an important difference between the new circuit and all the earlier ones, in which the leakage inductance and the resistance of the transformer windings introduce errors when the capacitance to earth from one terminal of the specimen changes. This effect may be thought of as poor regulation of the transformer when loaded, making the potentials of the two outer terminals of the winding no longer equal and opposite, so that equal capacitances on the two sides no longer produce a balanced condition. There is no corresponding effect in the circuit now proposed so long as the auxiliary arm of the network is left untouched.

(4) METHODS OF BALANCING THE LOSS COMPONENT

If the capacitance to be measured is associated with conductance, a second adjustment must be provided. In principle it may consist of variable resistance in series with C_v, or variable conductance in parallel with it. Neither is convenient in practice, however. It is difficult to introduce variable resistance so that it is in series with a measured capacitance without being also in series with some capacitance to earth. The only way to realize a variable conductance capable of suitably small changes would be to use a variable resistance of a few ohms in series with a large fixed resistance; such a variable conductance, or any form of variable resistance, would need double screening to prevent the appearance of stray capacitance to earth from intermediate points which are connected to neither oscillator nor detector. Since double-screened resistance and conductance boxes are not normally available, some more convenient device is desirable.

A suitable alternative is a T-network of resistors in the measuring arm, as shown in Fig. 5, where A is a negative

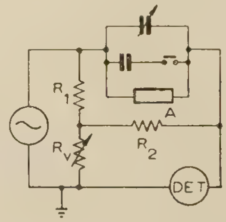


Fig. 5.—Resistive network for power-factor adjustment.

capacitance as in Fig. 1. Using the symbols shown in the diagram, the effective conductance across the specimen arm is approximately R_v/R_1R_2, which can easily be given values of the order of 10^{-6}–10^{-12} mho, and R_v is direct-reading in conductance. Since the fixed resistors can be of the high-stability carbon type, and therefore physically small, the capacitance appearing elsewhere than across the oscillator or the detector is small; and, since the variable resistance need be of only a few hundred ohms, the capacitance will be unimportant at frequencies up to about 1 Mc/s.

The alternative network shown in Fig. 6 has the advantage of

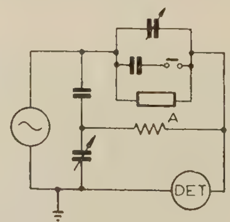


Fig. 6.—Network for power-factor adjustment, using a variable capacitor.

using variable capacitance instead of variable resistance. It is, however, more liable to introduce unwanted changes in the capacitance appearing in parallel with the specimen, when the effective conductance is changed.

Since the measuring arm of the network includes conductance as well as capacitance, the auxiliary arm also needs some conductance to allow an initial balance. The resistance of the transformer windings already provides it, and will probably allow measurements on specimens having a low power-factor; otherwise some more series or shunt resistance should be added,

and left in circuit for the measurements both with and without the unknown capacitor connected in the measuring arm of the network.

(5) STABILITY OF THE NETWORK

The main components used in the network will, of course, be of the types known to be stable enough for precise work. This need not always be true of the resistance network used to measure the power factor, because its admittance may be small compared with that of the capacitive arms of the network, and some instability in it is then tolerable. For example, for the measurement of a specimen having a power factor of the order of 0·001, the modulus of the admittance of the resistive network might be made about 0·005 of that of the capacitive arm, and then an instability 200 times as great as that of the capacitive components would be tolerable.

The component not normally made for precise work is the transformer. The balance conditions involve its turns ratio, its leakage inductance L, the resistance R and self-capacitance C_s of its windings, and the capacitance C_m between them. The ratio is not likely to change, but the other characteristics may do so. The transformer introduces unwanted admittances which can be evaluated by star-mesh transformations and have the approximate values given in Table 1.

For estimating the importance of the conductances, they may be divided by ω, so that they are comparable with capacitances. For typical transformers in which no unusual precautions have been taken, and which are suitable for a network in which $C = 20\,\mu\mu F$, values might be as given in Table 2.

If these terms are stable to within 1% of themselves, the whole network can be stable to 1 part in 10^6 at 1 kc/s, but only to 1 part in 10^4 at 1 Mc/s. With a transformer which is specially designed or thermostatically enclosed, the stability might be improved a little.

One remaining source of trouble is variation of the transformer characteristics with applied potential. This is minimized if the potential is kept always small compared with that which would saturate the core. In an experiment, the balance changed by 60 parts in 10^6 when the supply to the bridge was changed from 2 to 4 volts, using a transformer in which core saturation occurred with about 100 volts; in practice, such a supply can be kept constant within about 5%, thus allowing changes in balance of about 5 parts in 10^6.

Table 1

	Unwanted capacitance	Unwanted conductance
In parallel with the auxiliary capacitance C_k	$\omega^2 L(C_m + C_s + C_k)C_k$	$-\omega^2(C_m + C_s + C_k)RC_k$
In parallel with the specimen	$-\omega^2 LC_mC_k$	$-\omega^2 C_mRC_k$
Equivalent total in one arm	$\omega^2 L(2C_m + C_s + C_k)C_k$	$\omega^2(C_s + C_k)RC_k$

Table 2

	L	R	C_m	C_s	$\omega^2 L(2C_m + C_s + C_k)C_k$	$\omega(C_s + C_k)RC_k$
	μH	ohms	μμF	μμF		
For use at 1 kc/s	10 000	20	50	100	$90C_k \times 10^{-6}$	$15C_k \times 10^{-6}$
For use at 1 Mc/s, not screened	5	3	10	20	$12\,000C_k \times 10^{-6}$	$720C_k \times 10^{-6}$
For use at 1 Mc/s, screened	5	3	0	30	$10\,000C_k \times 10^{-6}$	$1\,100C_k \times 10^{-6}$

(6) COMPARISON OF PERFORMANCE OF BRIDGES

To test the new circuit, it was set up and balanced, and capacitance and conductance were then introduced in one of the positions where strays would appear in practice. The changes in balance were noted. The same experiment was carried out with an ordinary 'transformer bridge'. The transformer was the same for each circuit; the resistance of each winding connected to the bridge was 150 ohms, and the leakage inductance associated with it was 150 μH. The capacitance in the arms of the network was about 20 μμF and the frequency was 10 kc/s. From these data the error to be expected in the latter bridge can be calculated. The results are given in Table 3. The two bridges are alike in sensitivity and stability.

Table 3

	Error in ordinary 'transformer bridge'				Error in proposed network	
	Calculated		Observed			
	In C	In G	In C	In G	In C	In G
	μμF	micro-micromhos	μμF	micro-micromhos	μμF	micro-micromhos
With capacitance of 1 000 μμF added across one half of transformer winding	+0·010	+12 000	+0·012	+17 000	$+0{\cdot}000_5$	−60
With conductance of 24 micromhos added in same position	−0·075	+320	−0·078	+370	$<0{\cdot}000_5$	<10

(7) EFFECTS AT HIGH FREQUENCIES

At frequencies of the order of 1 Mc/s the impedance of the connecting leads between the specimen and the rest of the

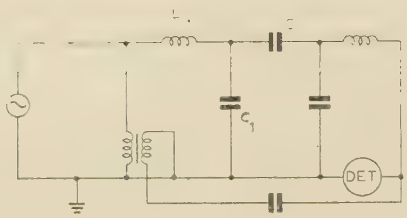


Fig. 7.—Bridge network with stray inductance and capacitance.

apparatus may not be negligible. Fig. 7 shows the impedances which cause trouble; it is sufficient to consider any one of them—say L_1. Capacitance to earth now appears from a point which is connected to neither oscillator nor detector, and it will cause errors, as can be seen by converting the star L_1, C_1, C to an equivalent mesh. If the network is being used to compare two nearly equal capacitances, the effect can be minimized by arranging the wiring in the form suggested by Fig. 8, and con-

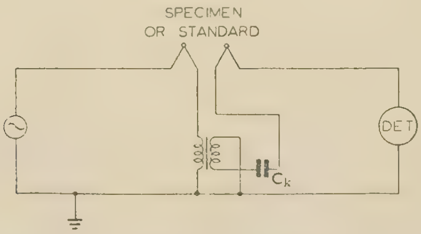


Fig. 8.—Arrangement for avoiding trouble from stray inductance.

necting the specimen and the standard successively in the position shown, which method was used in obtaining the results quoted in the companion paper.[1] Another plan, if it is known that only one of the earth-capacitances is varying (as can usually be arranged by screening), is to provide an additional adjustment in the network so as to keep C_1 constant. An extra detector is introduced, as shown at D_2 in Fig. 9 (or the original detector is

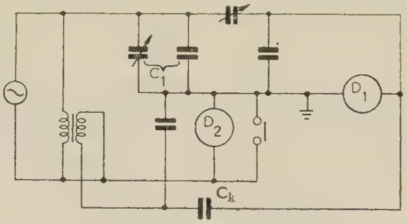


Fig. 9.—Method of adjusting network so that stray capacitance is constant.

switched to this position), and the capacitance C_1 is adjusted to produce a balance. The detector D_2 is then preferably short-circuited, and the normal balancing adjustment is made. After the substitution, both adjustments are repeated. This procedure is like that followed when using a Wagner earth. The difference is in the great rapidity of the convergence of the balances, so that repeated switching between the two is unnecessary. The Wagner earth, when it is used in a normal bridge, is probably important; the procedure just described is a refinement, and is needed only at high frequencies.

(8) CONCLUSIONS

The new measuring circuit has the following properties.

(*a*) The measurements of capacitance and conductance are practically unaffected by stray capacitance or conductance. In this respect the network is better by a factor of about 100 than one which is good by accepted standards.

(*b*) The bridge balance is almost independent of frequency.*

(*c*) There are normally no switching operations such as those needed for a Wagner earth.

(*d*) Both oscillator and detector have one terminal earthed.

No other network for measuring capacitance or resistance combines these advantages. On the other hand, the method can be used only where substitution is possible; capacitance can be measured only by comparison with capacitance, and resistance with resistance. There are, however, some exceptions to this limitation; for example, inductance can be measured in terms of capacitance by making the measuring arm a parallel-resonant circuit.

The circuit can in principle be used at any frequency for which a transformer can be made—say 4 c/s to 20 Mc/s—and it is particularly useful for measuring capacitances of less than 100 $\mu\mu$F, which are comparable with, or less than, the stray capacitances to earth. At frequencies up to about 1 Mc/s, it is suitable for work of the highest precision, yet it is so simple that it is worth using wherever the restriction to substitution methods can be accepted.

The circuit is the subject of a patent application.

(9) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the Post Office and to the Controller of H.M. Stationery Office for permission to publish this paper.

(10) REFERENCES

(1) Lynch, A. C.: 'A Method for the Precise Measurement of Permittivity of Sheet Specimens' (see page 359).
(2) Blumlein, A. D.: British Patent No. 323037.
(3) Starr, A. T.: 'A Note on Impedance Measurement', *Wireless Engineer*, 1932, **9**, p. 615.
(4) Trowbridge, A.: 'On the Differential Transformer', *Physical Review*, 1905, **20**, p. 65.
(5) Cole, R. H., and Gross, P. M.: 'A Wide-Range Capacitance-Conductance Bridge', *Review of Scientific Instruments*, 1949, **20**, p. 252.

DISCUSSION ON THE ABOVE TWO PAPERS BEFORE THE MEASUREMENT AND CONTROL SECTION, 29TH JANUARY, 1957

Mr. C. G. Garton: The transformer, or coupled-inductive-arm, bridge, of which the bridge described by the author is an improved version, has a history dating from the 1920's. It is remarkable that it has not come into more general use, since it has considerable advantages over more conventional bridges, including the Schering bridge. I have myself used it in one form, without, of course, the author's improved connection, and I described it incidentally in the course of an Institution paper on another subject.† The circuit could, without difficulty, be made sensitive to a loss angle of 20 microradians, and this was subsequently improved to 2 microradians for the purpose of checking loss-free capacitors. Extremely careful screening was, however, essential. The distinctive feature of the Lynch bridge, which so

* This claim, often made for bridge circuits, conceals a number of assumptions: it is as fair a claim for this circuit as for any other.

† Garton, C. G.: 'The Drying Process in Paper, as determined by Electrical Methods', *Journal I.E.E.*, 1940, **86**, p. 369.

ffectively reduces the necessary screening, is, of course, the ngenious connection of the supply circuit across one branch nly of the coupled inductive arms. With this radical improvement, the Lynch bridge can and should replace the Schering ridge for measurements at low voltages.

One limitation in the sensitivity of bridges arises from the fact hat, even if the in-phase balance is independent of frequency, he quadrature balance is not. A bridge balanced at the fundamental frequency is therefore out of balance for harmonics, if he test object possesses a loss angle. For accurate work, this isually necessitates elaborate filtering circuits. The transformer ridge has a great advantage over the Schering bridge in this espect. The latter introduces a quadrature component of the orm ωCR, so that, considering the measurement of loss angle of a dielectric, the bridge is balanced for all harmonics only if he loss angle of the sample for the nth harmonic is n times that or the fundamental. In fact, for low-loss dielectrics, where this question is of interest, the loss angle is substantially independent of frequency, so that the Schering bridge is out of balance, for the nth harmonic, by an amount $(n - 1) \tan \delta$. The bridge is usable only because the harmonic amplitude normally falls much aster than $1/n$.

The phase correction of the transformer bridge, on the other hand, is of the form $1/\omega CR$, so that the nth harmonic unbalance s given by $(1 - 1/n) \tan \delta$—a limited quantity which cannot exceed $\tan \delta$ for any frequency. It is therefore possible, without filter circuits, to use the Lynch bridge at a higher sensitivity than would be practicable with a Schering bridge.

Mr. J. K. Webb: When the author devised his non-contact electrode scheme he was thereby set the problem of balancing accurately a direct capacitance of the order of 10 pF associated with much larger and varying stray capacitances. He has solved it in a very elegant manner by means of what may, at first sight, appear to be only a trifling modification to a well-established method, but which nevertheless has most important consequences. On the assumption that the permittivity of polythene is required to the accuracy stated, I should be interested to learn, first, what is the standard deviation between various samples taken from the same batch, and secondly, whether any trouble has been experienced from the dimensional instability of samples.

Mr. Lynch (*in reply*): The dimensional instability of samples of dielectric is likely to be less serious when the method of electrical measurement does not require any pressure to be exerted on them, and I have not noticed any trouble of this kind. The method described in the papers is in use commercially for material whose permittivity is specified to within $\pm 0{\cdot}3\%$, but I have no information about variation between samples.

I agree with Mr. Garton's remarks about the effect of harmonic voltages in the Schering bridge, although I think the attenuation of the higher harmonics is mainly a question of transformer and amplifier responses rather than, as he implies, oscillator waveform.

621.317.733.029.3 : 621.314.7

Paper No. 2247 M
Dec. 1956

A SIMPLE TRANSFORMER BRIDGE FOR THE MEASUREMENT OF TRANSISTOR CHARACTERISTICS

By W. F. LOVERING, M.Sc., Associate Member, and D. B. BRITTEN, B.E., Graduate.

(*The paper was first received* 9*th May, and in revised form* 28*th September*, 1956. *It was published in December*, 1956, *and was read before the* MEASUREMENT AND CONTROL SECTION 29*th January*, 1957.)

SUMMARY

The paper describes a transformer bridge for the measurement of the impedance parameters of point-contact or junction transistors at 1 000 c/s. The bridge is simple and rapid in operation and measures both real and imaginary components of the parameters. A modification of the bridge connections measures the components of a simple equivalent circuit of a junction transistor.

Further modifications are suggested which will allow for the measurement of admittance parameters or h-parameters, direct measurement of the impedance parameters of junction transistors connected in the common-emitter circuit, or measurements at radio frequencies.

(1) INTRODUCTION

The characteristics of a three-terminal network may be specified by four independent parameters.. The sets of parameters which have been most widely used are the impedance parameters, the admittance parameters or the h-parameters.

Impedance parameters are defined by

$$V_1 = I_1Z_{11} + I_2Z_{12} \qquad (1)$$
$$V_2 = I_2Z_{22} + I_1Z_{21} \qquad (2)$$

Admittance parameters are defined by

$$I_1 = V_1Y_{11} + V_2Y_{12} \qquad (3)$$
$$I_2 = V_2Y_{22} + V_1Y_{21} \qquad (4)$$

h-parameters[1] are defined by

$$V_1 = I_1h_{11} + V_2h_{12} \qquad (5)$$
$$I_2 = I_1h_{21} + V_2h_{22} \qquad (6)$$

The voltage and current directions are shown in Fig. 1.

If one set of parameters is measured the other sets may be calculated;

e.g. $h_{11} = Z_{11} - Z_{12}Z_{21}/Z_{22}$ and $Z_{11} = Y_{22}/(Y_{11}Y_{22} - Y_{21}Y_{12})$

A set of parameters may be used to derive a simple equivalent circuit. Impedance parameters lead naturally to an equivalent star of three impedances and a generator; admittance parameters lead to an equivalent mesh of three impedances and a generator.

A network containing distributed resistance and capacitance in general has parameters which vary with frequency, and a simple four-element equivalent circuit is valid at one frequency only.

The small-signal characteristics of a transistor are those of a three-terminal network and may be specified by any convenient set of parameters—these, however, vary with frequency. The relative merits of these three sets of parameters have been discussed in the literature. Pritchard[2] concludes that the choice will be determined mainly by the manner in which the transistor is used in its associated network, but that ease of measurement is an important consideration.

Several methods have been described for the measurement of transistor characteristics; oscilloscope displays[3,4,5,6] plot the voltage/current characteristics but are of limited accuracy. Voltage and current measurements have been used to evaluate impedance parameters[7] or h-parameters.[1] Bridge methods have been used to measure impedance parameters,[8,9] admittance parameters,[10,11] h-parameters[12,13] or a mixture of h-parameters and impedance parameters.[14]

To measure impedance parameters it is necessary to make I_1 zero when $Z_{22} = V_2/I_2$ and $Z_{12} = V_1/I_2$, or I_2 zero when $Z_{11} = V_1/I_1$ and $Z_{21} = V_2/I_1$. Several workers have arranged to approximate to this condition by including a high resistance in series with Z_{11} or Z_{22}. This procedure is satisfactory with point-contact transistors for which Z_{22} is only a few thousand ohms, but is not possible with junction transistors having a value for Z_{22} of several megohms. Two circuits[8,14] have been described in which I_2 may be reduced to zero by a counter-balancing voltage, but a large resistance is used to simulate $I_1 = 0$.

The impedance parameters of a junction transistor differ greatly in magnitude. Z_{11} and Z_{12} may be a few hundred ohms, whilst Z_{22} and Z_{21} may be several megohms. Few bridges are equally well suited to the measurement of both high and low impedances, and for this reason the impedance parameters of point-contact transistors have been measured, but for junction transistors, admittance parameters or h-parameters have been measured.

A bridge using inductively coupled ratio arms[15] may be used to measure a few ohms or many megohms with little error, and such a bridge is very simply adapted to measure the impedance parameters of a three-terminal network.[16]

(2) THE BRIDGE CIRCUIT

The arrangement is illustrated in Fig. 1. The phase and magnitude of V_1 is adjusted so that I_2 is zero when

$$V_1 = I_1Z_{11} \text{ and } V_2 = I_1Z_{21}$$

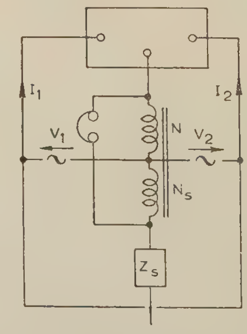


Fig. 1.—Basic bridge circuit.

Mr. Lovering was formerly in the School of Electrical Engineering, New South Wales University of Technology, Australia, and is now in the Electrical Engineering Department, Imperial College of Science and Technology, University of London.

Mr. Britten was formerly in the School of Electrical Engineering, New South Wales University of Technology, and is now in the Postmaster General's Department, New South Wales, Australia.

If Z_s is connected to V_2 and adjusted for zero signal in the detector,

$$Z_s = Z_{21}Ns/N \text{ and } Z_{11} = Z_{21}V_1/V_2$$

The calculation of Z_{11} from Z_{21} requires a knowledge of the vector ratio V_1/V_2. This is unnecessary if a second balance is taken with Z_s connected to V_1, when

$$Z_s' = Z_{11}Ns/N$$

If the circuit is modified to allow for the necessary d.c. biasing, the impedance parameters of a transistor may be measured in a similar fashion.

A merit of the transformer bridge is that it measures direct impedances even in the presence of relatively large earth capacitances.[15] All connections to the transistor may be made through screened wire, and earthed screens may be used between the three bridge terminals; stray capacitances may thus be eliminated.

The basic circuit of the bridge as modified for the measurement

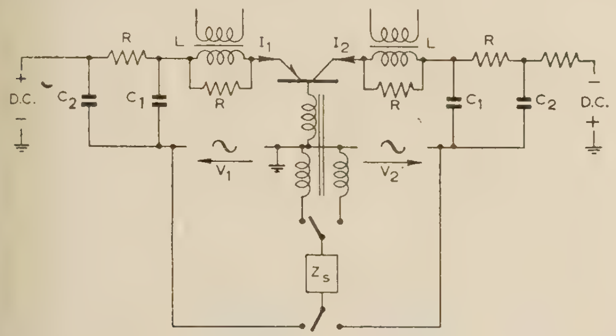


Fig. 2.—Bridge with current-detection circuits and d.c. biasing.

of transistor parameters is shown in Fig. 2. To detect the presence of the current I_1 or I_2 the primary winding of a transformer is included in each lead.

(2.1) Current Detection and D.C. Biasing

The d.c. by-passing and current-detection circuits add an impedance in series with both Z_{11} and Z_{22}. This impedance may be made to approximate to a pure resistance R at all frequencies above a few hundred cycles per second, by using a large capacitance (about 100 μF) for C_2, a capacitance of about 1 μF for C_1 and by making R equal to $\sqrt{(L/C_1)}$, where L is the inductance of the current-detection-transformer primary winding. The resistance R can be compensated by a resistance of appropriate value connected in series with Z_s.

In practice the impedance added in series with Z_{11} (or Z_{22}) has a small reactive component, but this is only a fraction of ωL. If ωL does not exceed 20 ohms the reactive term is a fraction of an ohm and is negligible.

(2.2) The Current-Detection Transformer

The current to be detected is of the order 10^{-8} amp when junction transistors are tested. This produces a secondary voltage of approximately 1·6 μV at 1000 c/s, if the transformer has a primary inductance of 0·6 mH and a secondary inductance of 1·0 henry. A sensitive amplifier and headphones suffice to detect this voltage.

The potential difference between the primary winding and earth may be several volts, and an appreciable voltage may be developed across the secondary winding due to the capacitance between the windings. Direct capacitance coupling can be reduced to small proportions by earthed screens surrounding each winding. A further improvement results from the use of a low-inductance secondary winding, with one side earthed, supplying a step-up transformer. The combination of earthed screens and a low-voltage secondary winding effectively eliminates all direct capacitance coupling between the current circuit and the detector.

The capacitance between the primary winding and screen, however, causes a capacitive current to flow in the winding and in the screen. If the earth point on the screen is incorrectly positioned the m.m.f.'s due to the current in the screen add to those due to the current in the windings, and an e.m.f. is induced in the secondary winding.

By correct positioning of the earth connection it is possible to reduce the effects of capacitive currents, but it is difficult to remove them completely. As a final measure, therefore, a

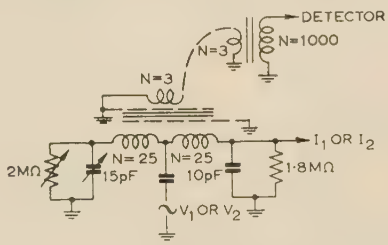


Fig. 3.—Current-detection circuit.
N = Number of turns.

balancing circuit is used as illustrated in Fig. 3. The variable capacitor and variable resistor are adjusted until there is no secondary voltage when a potential difference is applied between the winding and earth and the output lead is open-circuited. This arrangement makes it possible to compensate for the capacitance from the output lead to earth and a screened lead may be used.

(2.3) Attenuator and Phase-Shifter

In a transistor the ratios Z_{11}/Z_{21} and Z_{12}/Z_{22} are small, and the ratio V_1/V_2 required for current balance is also small. V_2 is supplied from an external source of about 5 volts, and V_1 is produced by causing a current, derived from V_2, to flow in a resistance R. The current is continuously variable by means of a potentiometer; a second potentiometer supplies quadrature current and allows for V_1 to be shifted in phase by a few degrees. A third potentiometer is used for fine control of V_1.

When Z_{11} or Z_{12} is being measured the standard impedance Z_s is connected in parallel with R. To avoid the necessity to readjust the attenuator when changing from the measurement of Z_{11} to that of Z_{21} (or from Z_{22} to Z_{12}), R is made much smaller than the lowest probable value of Z_s; a resistance of about one ohm is found to be quite satisfactory.

(2.4) Ratio Transformer

The transformer used has a ring core of spirally wound strip Permalloy, having a cross-sectional area of 1·6 cm^2 and a mean diameter of 5 cm. The ring is uniformly wound with 300 turns of No. 26 s.w.g. enamelled wire, with a second winding of 30 turns uniformly wound upon it. A third winding is made up of six equally spaced groups, each of three turns, connected in parallel.

The impedances to be measured are of the order of a few hundred ohms (Z_{11} or Z_{12}) or many thousand ohms (Z_{22} or Z_{21}). For the former a ratio of 30 : 300 is used and for the latter a ratio of 30 : 3, so that the standard resistor is used within a range 1 000–500 000 ohms.

At balance the leakage impedance of the transformer winding adds to Z_{12} and Z_{21}, but the 30-turn winding has a resistance of

only 0·16 ohm and a leakage inductance, with either of the other windings short-circuited, of less than 1 μH. Errors from this source are quite negligible.

Balance is detected by the absence of a voltage across the 300-turn winding; this winding also has a negligible leakage impedance so that zero voltage across it corresponds to equal and opposite m.m.f.'s in the two windings in use.

(2.5) The Complete Bridge Circuit

The circuit of Fig. 4 shows how the basic circuit is modified by the addition of switches, etc., to produce a practical laboratory instrument for measurements on transistors connected in the common-base arrangement.

A three-pole four-position switch is used to select the circuit arrangement and appropriate transformer ratio to measure Z_{11}, Z_{21}, Z_{12} or Z_{22}. A two-pole two-position switch is used to connect the amplifier either to the current-detection transformer or to the ratio transformer.

When the ratio transformer is unbalanced an impedance is added to Z_{12} and Z_{21}, and to avoid the necessity of repeated adjustments this added impedance should be made negligible. This is conveniently achieved by short-circuiting the 300-turn winding on the ratio transformer. The impedance added to Z_{12} and Z_{21} is then only the leakage impedance of the 30-turn winding and is negligible. The short-circuiting is done by the switch used to switch the amplifier.

Since some point-contact transistors are unstable with the emitter short-circuited a resistance of 1 500 ohms is added in the emitter lead in addition to the resistance of 100 ohms resulting from the current-detection circuit. A compensating resistor of 16 000 ohms is connected to the 300-turn winding of the ratio transformer and is in series with Z_s when Z_{11} is being measured. A similar compensating resistor of 10 ohms is added in series with the 3-turn winding for use when Z_{22} is measured; this refinement may not be justified since a resistance of 100 ohms in series with Z_{22} represents an error of less than 0·1% if Z_{22} exceeds 100 000 ohms. The added resistance in the emitter circuit reduces the sensitivity of the bridge, and, when possible, it is removed from the circuit.

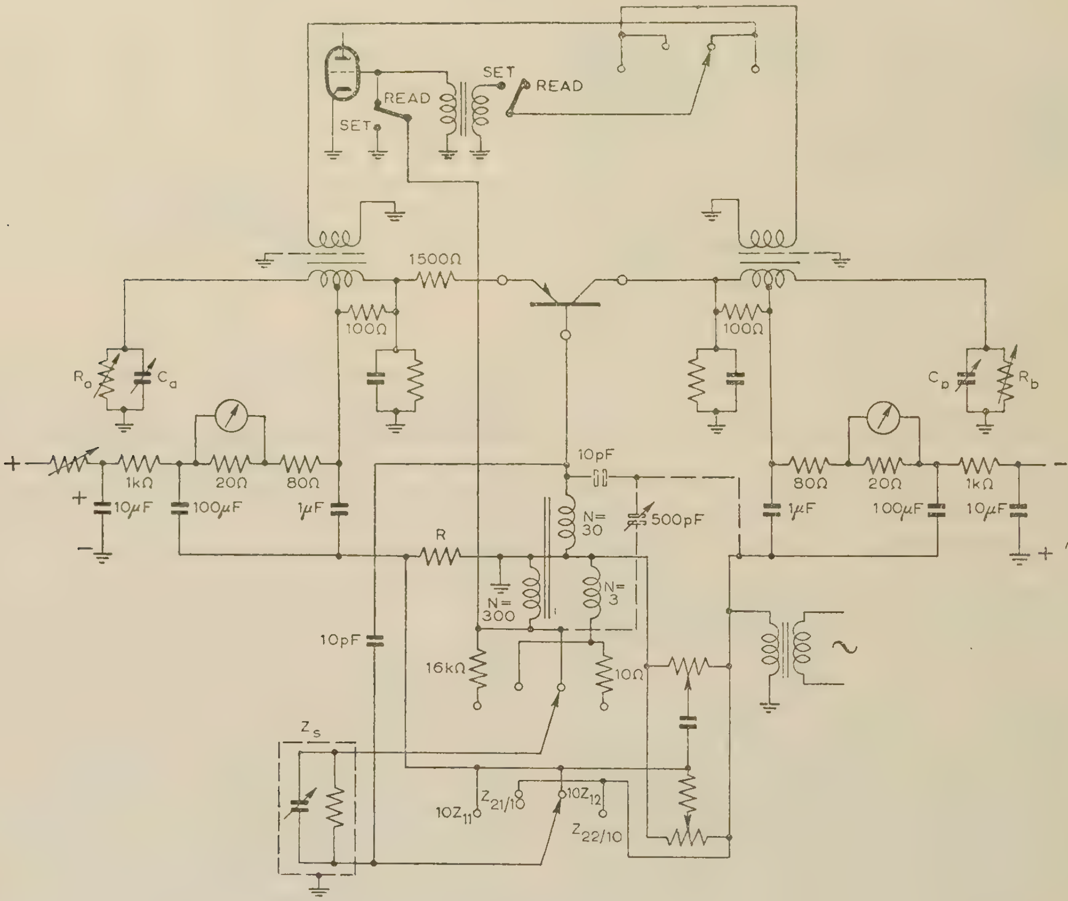


Fig. 4.—Complete bridge circuit.
Alternative balance capacitance connections are dotted.

For the standard arm a five-dial decade resistor is used in parallel with an appropriate capacitance. For measurements of Z_{11} or Z_{12} for junction transistors a decade capacitor is used, and for measurements of Z_{22} or Z_{21} a 10–500 pF variable air capacitor is used. The standard is connected to the bridge by screened leads.

Capacitances to earth do not affect the measurement, since at balance there is a negligible potential drop across the ratio transformer.

The decade capacitor may be dispensed with if the 500 pF variable capacitor is connected permanently to V_2 as shown by the dotted connection in Fig. 4. In this case the capacitance readings at balance correspond to $C_{11}(V_1/V_2)_{(I_2=0)}$, C_{21}, $C_{12}(V_1/V_2)_{(I_1=0)}$, or C_{22}. Thus in the Z_{11} position the equivalent parallel capacitance of Z_{11} is given by $C_{11} = C'_{11}R_{21}/R_{11}$, where C'_{11} is the measured capacitance allowing for the transformer ratio. Similarly $C_{12} = C'_{12}(R_{22}/R_{12})$.

Three spring connectors are provided for connecting the transistor under test. Each connector is surrounded by an earthed metal screen so that the bridge measures the transistor parameters without added capacitance.

The 10 pF capacitor connected from the selector switch to the common terminal of the transistor serves to balance part of the standard capacitance. In the Z_{22} or Z_{21} position the transformer ratio is 10 : 1 and the standard capacitance is 10–500 pF, which can thus balance a transistor capacitance having any value between −9 and +40 pF.

(2.6) Bridge Adjustments

The current-detection circuits are balanced by a preliminary adjustment. With no external connections to the three test terminals, the two-position switch is placed in the 'set' position and the semi-variable resistors and capacitors R_a, R_b, C_a and C_b are adjusted for zero signal. This adjustment compensates for the effect of unbalance in the current-detection transformer and for the effects of capacitance to earth from the test terminals and connecting wires. The adjustments, once made, are permanent and need not be repeated.

To measure transistor parameters the transistor is connected in circuit and the d.c. operating conditions are adjusted. The four-position switch is set to the appropriate position and the two-position switch placed in the 'set' position. The attenuator and phase-shift controls are then adjusted for zero signal.

The switch is then placed in the 'measure' position and the standard resistor and capacitor are adjusted for zero signal. The impedance is then given by Z_s times the appropriate ratio.

The selector switch selects the parameters in the order Z_{11}, Z_{21}, Z_{12} and Z_{22}. In the Z_{11} position the attenuator and phase-shifter are adjusted to reduce I_2 to zero. If the switch is now moved to the Z_{21} position I_2 remains zero and Z_{21} may be measured without further readjustments of the attenuator and phase-shifter. Measurement of a complete set of parameters thus requires only six balance adjustments, namely two settings of the attenuator and phase-shifter and four settings of the standard resistor and capacitor.

(3) RESULTS

(3.1) Bridge Performance

The noise level in the amplifier is such that an input signal of 1 μV is just detectable above the noise. The zero adjustment of I_1 or I_2 thus makes this current less than 10^{-8} amp. The ratio transformer develops 3 μV across the 300-turn winding for an out of balance of 3×10^{-7} AT; an out-of-balance current of 3×10^{-9} amp in the 30-turn winding is thus detectable.

Assuming a signal current of 3 μA in the transistor, the above figures indicate that the final balance should be adjusted to a precision of one part in 1 000 but that the resulting measurement is liable to an error of one part in 300 due to an imperfect initial current balance. Measurements of Z_{22} and Z_{12} for a star of precision resistors and capacitors confirmed the calculation.

Measurements on transistors are less accurate because of transistor noise. An equivalent noise current of 3×10^{-8} amp, for example, causes a signal current of much less than this value to be undetectable in either balance. With a signal current of 3 μA the precision of the final balance becomes one part in 100, and a further error of one part in 100 results from the imperfect initial current balance.

The bridge is very easy to use; a complete set of parameters can be obtained in two or three minutes. Point-contact transistors require no capacitance measurements, but the precision of adjustment is limited by the relatively high noise level. Z_{11} may be measured to within ±10%; the remaining parameters to within ±3%. The 1 500-ohm resistance in the emitter lead was found adequate to ensure stability with all the point-contact transistors tested.

Junction-transistor resistance parameters may be measured to within ±1% using a collector-circuit signal of less than 5 volts, or a signal current of less than 4 μA. The capacitances associated with Z_{22} or Z_{21} may be measured to within ±2 pF, and those associated with Z_{11} or Z_{12} to within ±6%.

An advantage of the bridge is that non-linear operation of the transistor produces harmonics which cannot be balanced out. A good balance is thus a direct indication of linear operation.

(3.2) Parameters of Point-Contact Transistors

The variation of the impedance parameters with emitter current for two point-contact transistors of different make is illustrated in Fig. 5. All the parameters were found to be purely resistive.

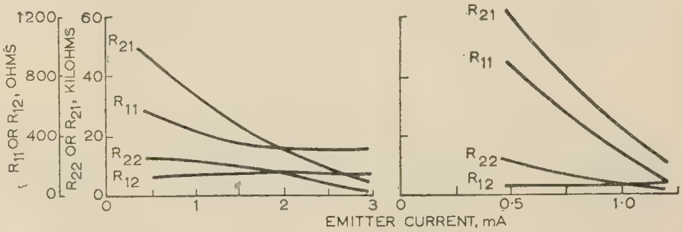


Fig. 5.—Parameters of two makes of point-contact transistor.

One of the transistors was very noisy, and the accuracy of measurement of Z_{11} was very poor (±20%).

The measurement does not give a direct indication of the current gain, but this is given very closely by R_{21}/R_{22}. For point-contact transistors, where α is usually greater than 2·0, this calculation is reasonably accurate.

(3.3) Junction-Transistor Parameters

The impedance parameters of one junction transistor at 1 000 c/s as a function of emitter current are illustrated in Fig. 6.

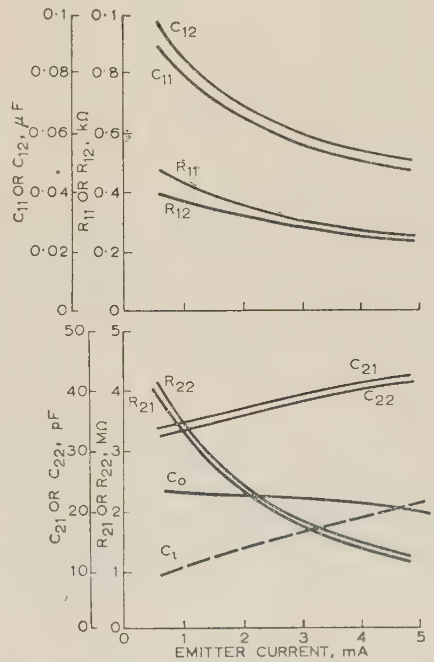


Fig. 6.—Parameters of a junction transistor.
C_0 and C_t refer to the equivalent circuit of Fig. 7.

In all cases these are capacitive, and each impedance is specified as a resistance in parallel with a capacitance; thus V_{12} and C_{12} in parallel represent Z_{12}.

This presentation has been adopted for convenience since the parallel combination is that which is measured.

As an example the values corresponding to an emitter current of 3 mA are as follows:

$R_{11} = 312$ ohms $\quad C_{11} = 0{\cdot}052\,\mu F$
$R_{12} = 295$ ohms $\quad C_{12} = 0{\cdot}058\,\mu F$
$R_{21} = 1{\cdot}61$ megohms $\quad C_{21} = 38$ pF
$R_{22} = 1{\cdot}70$ megohms $\quad C_{22} = 36$ pF

An equivalent circuit having these parameters is illustrated in Fig. 7(*a*).

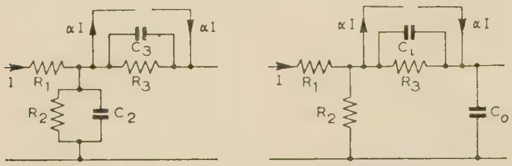


Fig. 7.—Alternative equivalent circuits of a junction transistor with emitter current of 3 mA.

$R_1 = 17$ ohms $\quad R_2 = 295$ ohms $\quad R_3 = 1{\cdot}7$ megohms $\quad \alpha = 0{\cdot}95_1$
$C_2 = 0{\cdot}058\,\mu F \quad C_3 = 36$ pF $\quad C_i = 26$ pF $\quad C_0 = 10$ pF

The resistances in the equivalent T-circuit may be considered as the lumped emitter, base and collector resistances, but the capacitances may not be assumed to represent the simple physical capacitances of the transistor.

It may be shown that a transistor having a collector-to-base capacitance C_0, an internal collector capacitance C_i and a current amplification factor $\alpha_0/(1 + j \tan kf)$ has impedance parameters given approximately by

$$1/Z_{12} = 1/R_{12} + j\omega C_0 R_{22}/R_{12}$$
$$1/Z_{11} = 1/R_{11} + j\omega C_0 R_{12} R_{22}/R_{11}^2$$
$$1/Z_{22} = 1/R_{22} + j\omega(C_0 + C_i)$$
$$1/Z_{21} = 1/R_{21} + j\omega(C_0 + C_i)R_{22}/R_{21} + j\, tan\, kf/R_{21}$$

if $R_{21} \gg R_{12}$ and if $\omega(C_0 + C_i)R_{12} \ll 1$ and $\tan kf \ll 1$.

The values of C_0, C_i and $\tan kf$ may be deduced from the measured values of C_{11}, C_{12}, C_{21} and C_{22}.

Applying these equations to the parameters of the previous example gives the equivalent circuit of Fig. 7(*b*). Such a circuit is in many ways more convenient, but it may not be considered as an exact equivalent of the transistor at any frequency other than that of measurement. It is interesting to note that the modified bridge connection (shown dotted in Fig. 4) measures C_0 directly. With this connection the measured capacitance C'_{12} is equal to $C_{12}R_{12}/R_{22}$ or $C'_{12} = C_0$. The equivalent circuit may thus be derived rapidly:

$$C_0 = C'_{12},\ C_i = C_{22} - C'_{12},\ \tan kf = \omega(C_{21}R_{21} - C_{22}R_{22})$$

and $\alpha_0 = R_{21}/R_{22}$.

The values of C_0 and C_i deduced from the measurements are shown plotted in the curves of Fig. 6. At the frequency of 1 000 c/s the minimum value of tan kf which could be detected with certainty was about 0·02; this corresponds to a cut-off frequency, $f_{\alpha co}$, for the transistor of only 20 kc/s. The transistor tested had $f_{\alpha co} \simeq 200$ kc/s and no significant difference was found between $C_{21}R_{21}$ and $C_{22}R_{22}$.

(4) CONCLUSIONS

It has been shown that there is little difficulty in the measurement of the impedance parameters of point-contact or junction transistors connected in the common-base circuit and that a bridge with inductively coupled ratio arms is readily adapted for this purpose.

An arrangement of a screened transformer has been described which makes possible the detection of an alternating current of less than 10^{-8} amp in the presence of a large voltage between the transformer windings.

Measurements on a junction transistor have shown that the imaginary components of the impedance parameters may be readily measured at 1 000 c/s. The measurements show large capacitances associated with Z_{12} and Z_{11}. An equivalent circuit has been deduced having small capacitances in the collector circuit only. It has been shown that the components of this equivalent circuit may be readily measured by a simple modification of the bridge circuit.

The effects of current-amplification cut-off are not detectable at 1 000 c/s, but with a suitable detector the bridge could be used at frequencies up to 50 kc/s. At such a frequency the imaginary components could be measured with greater precision, and $f_{\alpha co}$ should be readily deduced from the measurements.

Unity-ratio-transformer bridges have been used at frequencies up to 100 Mc/s. It seems probable that a bridge of the type described in the paper, but using ferrite cores, could be made for a frequency of 1 Mc/s or even higher.

For rapid comparison of different makes of transistor the impedance parameters are quite suitable. The common-base circuit is convenient because all the parameters are positive; the current amplification factor, α, is calculated from Z_{21}/Z_{22}.

The bridge may be used for the determination of h-parameters, three of which may be derived simply from the impedance parameters

$$h_{12} = Z_{12}/Z_{22},\ h_{22} = 1/Z_{22},\ h_{21} = -Z_{21}/Z_{22}$$

The calculation of h_{11} from impedance parameters is laborious and inaccurate ($h_{11} = Z_{11} - Z_{12}Z_{21}/Z_{22}$), but this parameter may be measured directly by a simple modification. The a.c. supply V_2 is disconnected from the collector circuit, and the collector is short-circuited to the base by a large capacitance. The impedance h_{11} may then be measured with the selector switch in the Z_{11} or Z_{12} position.

The performance of many circuits depends upon the magnitude of $(1 - \alpha)$, which cannot be calculated accurately when α approaches unity. If the transistor is measured in the common-emitter connection, $-Z_{22}/Z_{21}$ approximates closely to $1 - \alpha$. In this arrangement Z_{12}, Z_{22} and Z_{11} are all positive, but Z_{21} is negative. Measurement of a negative impedance with a transformer bridge requires merely a reversal of the transformer winding connected to Z_s.

Measurements of transistor parameters in the common-

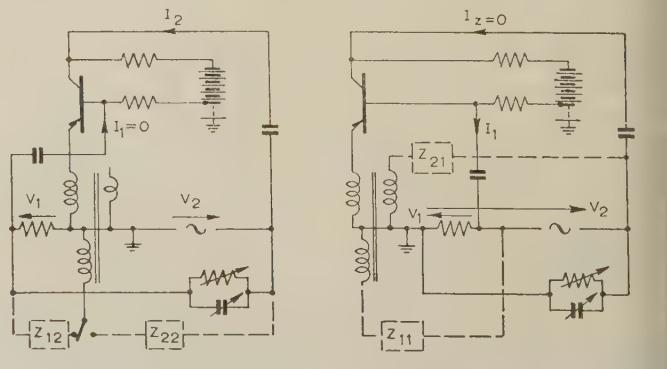


Fig. 8.—Connections required for measurements of parameters in the common-emitter circuit.

emitter circuit would require changes in the d.c. biasing, provision for reversing the connections to V_1, and provision for reversing one transformer winding or the addition of an extra three-turn winding. The appropriate circuits are illustrated in Fig. 8; unfortunately the departure of one of the authors halted this development.

(5) ACKNOWLEDGMENTS

The work was carried out in the School of Electrical Engineering in the New South Wales University of Technology, and thanks are due to Prof. R. F. Vowels for the use of facilities. The development of the bridge was commenced as a thesis project for the degree of Bachelor of Engineering. One of the authors (D. B. B.) wishes to acknowledge his debt to the Postmaster-General's Department in Sydney for part-time leave to continue the work.

(6) REFERENCES

(1) Knight, G., Johnson, R. A., and Holt, K. B.: 'Measurement of the Small Signal Parameters of Transistors', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 983.
(2) Pritchard, R. L.: 'Small Signal Parameters of Transistors', *Electrical Engineering*, 1954, 73, p. 902
(3) O'Neill, B. L., and Gutterman, A.: 'Methods and Equipment for Transistor Testing', *Electronics*, **26**, p. 144.
(4) Morrow, W. E.: 'Measurement of Transistor Parameters by C.R.O. and Other Means', *Proceedings of the Institute of Radio Engineers Convention Record*, 1953, Part 9, p. 45.
(5) Fewer, D. R.: 'Transistor Characteristics obtained by Pulse Techniques', *ibid.* 1953, Part 9, p. 53.
(6) Mathis, V. P., and Shaffner, J. S.: 'Quick Evaluation of Junction Transistor Characteristics by Oscilloscope Display', *ibid.*, 1953, Part 9, p. 72.
(7) Lehovec, K.: 'Testing Transistors', *Electronics*, 1949, **22**, p. 88.
(8) Cooke-Yarborough, E. H., Florida, C. D., and Stephen, J. H.: 'The Measurement of the Small-Signal Characteristics of Transistors', *Proceedings I.E.E.*, Paper No. 1614 R, February, 1954 (**101**, Part III, p. 288).
(9) Boothroyd, A. R., and Datta, S. K.: 'A Bridge for Measuring the A.C. Parameters of Point-Contact Transistors', *ibid.*, Paper No. 1646 M, March, 1954 (**101**, Part III, p. 294).
(10) Giacolleto, L. J.: 'Equipment for Measuring Junction Transistors over a Wide Range of Frequencies', *RCA Review*, 1953, **14**, p. 269.
(11) Giacolleto, L. J.: 'Bridges Measure Transistor Parameters', *Electronics*, 1953, **22**, p. 88.
(12) Boothroyd, A. R., and Almond, J.: 'A Bridge for Measuring the A.C. Parameters of Junction Transistors', *Proceedings I.E.E.*, Paper No. 1681, September, 1954 (**101**, Part III, p. 314).
(13) Coffey, W. N.: 'Measuring R.F. Parameters of Junction Transistors', *Electronics*, 1956, **29**, p. 153.
(14) Cooper, B. F. C.: 'A Bridge for Measuring Audio Frequency Transistors Parameters', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 796.
(15) Clark, H. A. M., and Vanderlyn, P. B.: 'Double-Ratio A.C. Bridges with Inductively-Coupled Ratio Arms', *Proceedings I.E.E.*, Paper No. 742 M. January, 1949 (**96**, Part II, p. 365).
(16) Lovering, W. F.: 'A New Impedance Bridge', *Journal I.E.E.* (New Series), 1955, **1**, p. 710.

DISCUSSION BEFORE THE MEASUREMENT AND CONTROL SECTION, 29TH JANUARY, 1957

Dr. R. F. Saxe: The paper represents an advance on previous work and is a good application of the transformer bridge, but the authors give two references to work where the current I_1 is made effectively zero by the inclusion of a high series resistance so that the impedance parameters can be measured. This appears to be not too serious a drawback if the input impedance of the transistor is low, so that there is a reasonable resistance to make I_1 effectively zero. However, I do not think the authors bring out clearly enough the advantages of this bridge over the two mentioned in References 8 and 14. The accuracy claimed seems to be about the same as that given in Reference 8.

In the form given in the paper the cut-off frequency of the bridge is unmeasurable if it is above 20 kc/s. The authors give a modification to cover this, but it has apparently not yet been done.

An advantage of the bridge is that non-linear operation becomes apparent, because a balance is not achieved. This appears to be a feature of one or two other bridges, and the authors give no indication of how near non-linear operation one is when operating under the conditions set out in the paper.

Cathode-ray-oscillograph methods of detection have been used before, and it is claimed that they give considerable discrimination against noise. I think that the aural method is probably not quite so good, and that a modification might be useful. The authors mention frequencies of 30 kc/s, which are, of course, beyond the range of most human ears.

The use of one or two components in Fig. 4 is not immediately apparent, e.g. the two 100-ohm resistors across the current-detection transformers. It is stated that a certain result is obtained from the current-detection circuit, but there is no indication as to how it is done.

Mr. L. G. Cripps (*communicated*): The bridge technique is an excellent way of measuring the Z-parameters of transistors, but the paper gives the erroneous impression that the measurements will lead to a useful equivalent circuit. This can be seen by comparing the two circuits of Fig. 7, which, although agreeing with the measured Z-parameters, do not agree with each other. (For example, consider the input impedance with the output short-circuited in the two cases.) It is thus misleading to present the results in the form of an equivalent circuit; a direct statement of the values of Z_{11}, Z_{12}, etc., is more useful.

The proposed arrangements of Fig. 8 are, from the d.c. point of view, common-emitter circuits, without stabilization of the working point against drift due to temperature. The drift can be large enough to cause difficulties when making accurate measurements. The problem may be dealt with by arranging that the transistor is in the common-base form from the d.c. point of view, whatever the a.c. circuit.

Messrs. W. F. Lovering and **D. B. Britten** (*in reply*): In reply to Dr. Saxe, the transformer bridge described in the paper has the merits of simplicity and ease of operation; transformer bridges are inherently free from errors due to stray capacitances, and a very wide range of impedance measurement is possible. The bridge described may be used for measurements on either point-contact or junction transistors. The measurements described in the paper were made with a collector voltage of about 16 volts and an a.c. signal in the collector circuit of about 5 volts; under these conditions non-linearity was unnoticed.

We prefer aural detection where possible, as the ear can readily pick out an a.f. note even in the presence of noise; however, cathode-ray detection and/or tuned amplifiers may be used at any frequency.

We are indebted to Mr. Cripps for criticism of the equivalent circuits of Fig. 7. These are approximately equivalent within the limits of error of the capacitance measurements, and the circuits illustrate the point that a large value of C_{12} may be due, in part at least, to a small capacitance from collector to base. The input impedance with the collector short-circuited is $Z_{11} - Z_{12}Z_{21}/Z_{22}$, and small errors in the measurement of the impedance parameters can give a considerable error in the calculated input impedance.

The bridge is very useful for the rapid comparison of transistors of all types, but the measurements must be supplemented with others to give a complete equivalent circuit representing a transistor. We hope, at some time, to see a published description of the circuits used by Mr. Cripps; we understand from a private communication that he includes, among others, measurements upon one form of transformer bridge.

621.385.2:621.3.011.4

Paper No. 2333 R
July 1957

THE CAPACITANCE BETWEEN DIODE ELECTRODES IN THE PRESENCE OF SPACE CHARGES

By C. S. BULL, Ph.D., Associate Member.

(*The paper was first received* 2nd *August, and in revised form* 19th *December*, 1956.)

SUMMARY

The capacitance between the electrodes of a planar condenser with no space charge can be calculated by finding the rate of change of charge on either electrode with the potential between them. In the presence of space charges, however, the result is different if the charge on the cathode is considered rather than the charge on the anode. The method is therefore not valid in this application. Accordingly, taking Child's approximation of zero emission velocities, the matter is reconsidered to take into account the charges which leave or accumulate between the electrodes when the anode potential is changed. It is then found that the capacitance of the active area of a space-charge-limited diode is zero, while that of a saturated diode varies from twice the cold capacitance to the cold capacitance as the anode voltage is changed from the saturation voltage to very high values at which space charge is negligible. Another feature of the results obtained is the sudden transition from high to low capacitance as the anode potential varies.

These results have a bearing on the efficiency and design of klystrons, affect the theoretical position in regard to the fluctuations of diodes, and reveal errors in the equations hitherto commonly used in the study of electron-inertia effects.

(1) INTRODUCTION

It has long been recognized that the capacitance between the electrodes of a valve is influenced by the presence of space charges. Since a valve is a non-linear device, conducting only in one direction, it has hitherto been customary to modify the relation $C = Q/V$ applicable to planar condensers with no space charge to

$$C = dQ/dV \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

which should be applicable to the diode.

To use this expression, Q, the charge on one electrode, the anode,[1] is estimated from the electric field at the anode, $(\partial V/\partial x)_A$, using Gauss's theorem. In this way the capacitance estimated from the charge on the anode is

$$C_A = \frac{\partial}{\partial V}\left[\epsilon_0\left(\frac{\partial V}{\partial x}\right)_A\right] \quad . \quad . \quad . \quad . \quad . \quad (2)$$

It is pertinent to ask whether the same capacitance would be obtained by considering the charge on the cathode, using an equation analogous to eqn. (2), i.e.

$$C_C = \frac{\partial}{\partial V}\left[\epsilon_0\left(\frac{\partial V}{\partial x}\right)_C\right] \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Physically we must expect both expressions, if correct, to give the same result. However, we see at once that, in general, $(\partial V/\partial x)_A \neq (\partial V/\partial x)_C$, and it would only be coincidental if the differentials with respect to V were identical.

Further considerations are probably best made under the approximation used by Child, namely that the emission velocity of the electrons is zero: Section 6 shows why this is necessary.

Using this approximation we can readily see, by elementary means, that, in the space-charge-limited range of the characteristic, $(\partial V/\partial x)_A$ is four-thirds of the field in the absence of space charges, so that

$$C_A = \tfrac{4}{3}C \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where C is the capacitance in the absence of space charge.

Also, since $(\partial V/\partial x)_C \equiv 0$

$$C_C \equiv 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

We thus immediately find a strong contradiction which arises from the neglect of the changes in charge density which take place in the space charge. The proper consideration of this effect differs from that of ordinary capacitances chiefly because the total space charge changes with the operating conditions. The valve therefore acts as a source or sink for electrons when non-stationary voltages are applied. Other factors, such as the electron emission energies, the flow of electrons in several directions in the same regions and the non-linear characteristic, merely complicate the matter, and we shall reduce these complications as far as possible by making appropriate approximations in our examination of the planar diode.

It will become apparent later that the capacitance can have values quite different from the 'cold' capacitance. Most strikingly, perhaps, it will be found that the capacitance under space-charge-limited conditions is zero. Physically this means only that, as the applied voltage changes, the current taken by the anode or cathode is at all instants equal to the steady current appropriate to the applied voltage, there being no capacitive component proportional to the rate of change of voltage. The movement of charges necessary to produce the changing electric fields and potentials at the anode is derived entirely by the space charge acting as a source or sink for electrons.

Since we are here concerned with the space charge as a macroscopic phenomenon, we shall speak of the flow of positive and negative charges, so adhering to ordinary electrostatic terminology to avoid excessive circumlocution.

We now proceed to study in turn each of the regimes under which the diode may be operated, and will discuss the implications of the results in Section 6.

(2) NEGATIVE ANODE VOLTAGES

Here no electrons leave the cathode, there is consequently no space charge, and therefore

$$C_A = C_C = C \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

(3) POSITIVE ANODE VOLTAGES, BUT WITH THE EMISSION SATURATED

Here we must take into account the fact that the potential is everywhere lowered by the presence of electrons, the extent depending on the anode voltage. We are concerned with how the space charge changes with, and how the electron motion is affected by, anode voltage.

In Fig. 1, C and A represent the cathode and anode planes. The points B and B′ represent two slightly different anode voltages, V and $V + \delta V$. The line OO′ represents zero potential.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Bull is Lecturer in Physics at the College of Technology, Birmingham.

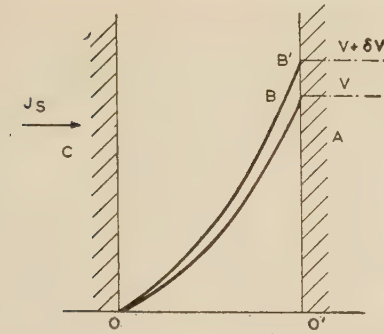


Fig. 1.—Change of potential between cathode and anode for a saturated valve at anode voltages of V and $V + \delta V$.

J_s = Total emission current density.

The lines OB and OB′ represent the course of the potential between cathode and anode for the anode voltages V and $V + \delta V$ respectively.

Since the anode current density, J, is a constant and equal to the total emission current density J_S, we can say that, since the space potential OB′ for $V + \delta V$ is everywhere higher than the potential distribution OB for anode voltage V, the space-charge density is everywhere lower. Furthermore, once an electron has started on its journey from cathode to anode, it cannot turn back, even though the anode voltage be reduced slightly.

Therefore, if we change the anode voltage from V to $V + \delta V$, by adding positive charge to the anode, the space charge during a time comparable with the time of flight must loose some electrons to the anode over and above those necessary to carry the current $J = J_S$. This charge will neutralize some of the positive charge added to the anode, so that to maintain the increased voltage a charge must be added larger than that necessary merely to establish the new electric field strength at the anode. It appears therefore that eqn. (2) will underestimate the capacitance.

At the cathode, on the other hand, the current cannot change, being equal to J_S all the time. The field, however, varies more rapidly than it would in the absence of space charges, from nearly zero when the anode voltage is just above the saturation voltage V_S, to V/x when the anode voltage is very high and space charge is consequently negligible. Eqn. (3) will therefore give a correct estimate of the capacitance, and we should modify eqn. (2) by adding a term representing the change in the total space charge, Q_S, getting

$$C_A = \epsilon_0 \frac{\partial}{\partial V}\left(\frac{\partial V}{\partial x}\right)_A - \frac{\partial}{\partial V} Q_S \quad . \quad . \quad . \quad . \quad (7)$$

Now we know from Poisson's equation that

$$\epsilon_0\left[\left(\frac{\partial V}{\partial x}\right)_A - \left(\frac{\partial V}{\partial x}\right)_C\right] = Q_S \quad . \quad . \quad . \quad (8)$$

Therefore
$$\epsilon_0\left(\frac{\partial V}{\partial x}\right)_A - Q_S = \epsilon_0\left(\frac{\partial V}{\partial x}\right)_C \quad . \quad . \quad . \quad . \quad (9)$$

From eqns. (3), (7), (8), and (9)

$$C_A = \epsilon \frac{\partial}{\partial V}\left(\frac{\partial V}{\partial x}\right)_A - \frac{\partial}{\partial V} Q_S = \epsilon_0 \frac{\partial}{\partial V}\left(\frac{\partial V}{\partial x}\right)_C = C_C \quad (10)$$

by which we obtain the same estimate of the capacitance by considering correctly the electric field and the processes taking place at either electrode and in the space charge.

The effect is not a time-of-flight effect, and is, in fact, effective at the lowest frequencies. At very high frequencies the flow of charges, $(\partial Q_S/\partial V)\delta V$, in and out of the space charge will not be simply calculated, and will produce negative-resistance effects which have been described previously.[2]

Formally, eqn. (10) is a complete expression, but a more detailed consideration of the values of Q_S, $(\partial V/\partial x)_C$ and $(\partial V/\partial x)_A$ would enable us to plot a curve of the variation of C_A (or C_C) with anode voltage above V_S.

To do this we follow Child's method of integrating Poisson's equation, but not taking $(\partial V/\partial x)_C = 0$, as follows:

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\rho}{\epsilon_0} = \frac{J_S}{\epsilon_0\sqrt{\left(\frac{2e}{m}\right)}} V^{-1/2}$$

Therefore
$$p\,dp = \frac{J_s}{\epsilon_0}\sqrt{\left(\frac{m}{2e}\right)} V^{-1/2} dV$$

where
$$p = \frac{\partial V}{\partial x}$$

and thus
$$p^2 = \frac{4J_S}{\epsilon_0}\sqrt{\left(\frac{m}{2e}\right)} V^{1/2} + \text{constant}$$

But, at
$$V = 0,\ p^2 = p_C^2 = \left(\frac{\partial V}{\partial x}\right)_C^2$$

Therefore
$$\left(\frac{\partial V}{\partial x}\right)^2 = \frac{4J_S}{\epsilon_0}\sqrt{\left(\frac{m}{2e}\right)} V^{1/2} + \left(\frac{\partial V}{\partial x}\right)_C^2$$

and
$$\left(\frac{\partial V}{\partial x}\right)_A^2 - \left(\frac{\partial V}{\partial x}\right)_C^2 = \frac{4J_s}{\epsilon_0}\sqrt{\left(\frac{m}{2e}\right)} V^{1/2}$$

where V is now the anode voltage.

But, from Poisson's equation,

$$\epsilon_0\left(\frac{\partial V}{\partial x}\right)_A - \epsilon_0\left(\frac{\partial V}{\partial x}\right)_C = Q_s$$

Therefore
$$\left[\left(\frac{\partial V}{\partial x}\right)_A + \left(\frac{\partial V}{\partial x}\right)_C\right] Q_S = 4J_s\sqrt{\left(\frac{m}{2e}\right)} V^{1/2}$$

We can now replace J_s by Child's expression relating the anode voltage at which saturation takes place and the anode-cathode distance, x,

$$J_S = \frac{4}{9}\epsilon_0\sqrt{\left(\frac{2e}{m}\right)} V_S^{3/2}/x^2$$

which gives

$$\left[\left(\frac{\partial V}{\partial x}\right)_A + \left(\frac{\partial V}{\partial x}\right)_C\right] Q_S = \frac{16}{9}\epsilon_0 V_s^{3/2} V^{1/2} x^{-2}$$

Therefore
$$Q_S = \frac{16}{9}\frac{\epsilon_0}{x^2} V_S^{3/2} \frac{V^{1/2}}{\left[\left(\frac{\partial V}{\partial x}\right)_A + \left(\frac{\partial V}{\partial x}\right)_C\right]} \quad . \quad . \quad (11)$$

This equation enables us to study two particular situations of interest, namely

(*a*) When V is just a little above V_s.
(*b*) When $V \gg V_s$.

In case (*a*), $(\partial V/\partial x)_C \simeq 0$ and $(\partial V/\partial x)_A \simeq \frac{4}{3}\frac{V}{x}$ so that

$$Q_s = \frac{16}{9}\frac{\epsilon_0}{x^2} V_S^{3/2} V^{1/2} \frac{3}{4}\frac{x}{V}$$

$$= \frac{4}{3}\frac{\epsilon_0}{x} V_S^{3/2} V^{-1/2}$$

and therefore $$\partial Q_s/\partial V = -\frac{2}{3}\frac{\epsilon_0}{x}\frac{V_S^{3/2}}{V^{3/2}}$$

$$\simeq -\frac{2}{3}\frac{\epsilon_0}{x}$$

Therefore, when $V \simeq V_s$,

$$C_C = C_A = \epsilon_0 \frac{\partial}{\partial V}\left(\frac{\partial V}{\partial x}\right)_A - \frac{\partial Q_s}{\partial V}$$

$$\simeq \epsilon_0 \frac{\partial}{\partial V}\left(\frac{4}{3}\frac{V}{x}\right) + \frac{2}{3}\frac{\epsilon_0}{x}$$

$$\simeq 2\frac{\epsilon_0}{x}$$

But $\epsilon_0/x = C$, the cold capacitance, so that when V is near but slightly greater than V_s, the capacitance of the valve is doubled. When $V \gg V_s$

$$(\partial V/\partial x)_C \simeq (\partial V/\partial x)_A \simeq V/x$$

and so $$Q_s \simeq \frac{16}{9}\frac{\epsilon_0}{x^2}V_S^{3/2}V^{1/2}\frac{x}{2V}$$

$$\simeq \frac{8}{9}\frac{\epsilon_0}{x}\frac{V_S^{3/2}}{V^{1/2}}$$

which tends to zero as $V \to \infty$. Consequently, $C_A = C_C \to C$ as V increases so that $V \gg V_S$.

(4) THE SPACE-CHARGE-LIMITED REGION

In the space-charge-limited region, $0 < V < V_S$,

$$Q_s = \epsilon_0\left(\frac{\partial V}{\partial x}\right)_A = \frac{4}{3}\epsilon_0\frac{V}{x} \quad . \quad . \quad . \quad . \quad (12)$$

Then $$\frac{\partial Q_S}{\partial V} = +\frac{4}{3}\frac{\epsilon_0}{x}$$

and $$C_A = \epsilon_0\frac{\partial}{\partial V}\left(\frac{4}{3}\frac{V}{x}\right) - \frac{4}{3}\frac{\epsilon_0}{x} \equiv 0$$

We have already observed that $C_C = 0$, so that $C_A = C_C = 0$.

(5) SUMMARY OF THE RESULTS FOR ZERO EMISSION VELOCITIES

We can collect the results obtained in Sections 2, 3, and 4 into three graphs, showing the anode current, the capacitance between the active faces of the electrodes, and the changes in capacitance, ΔC, as the anode voltage is varied. (This last is the only capacitance ordinarily accessible to measurement.) Fig. 2 shows the three graphs to the same scale of anode voltage. For all values of J_S the capacitance changes are identical, only the scale of V and V_S changing.

These results are quite different from those of previous investigations. It has hitherto been considered that the capacitance in the presence of space charge is either four-thirds or three-fifths of the cold capacitance, according to the authority. A striking feature of the conclusions is the suddenness of the transitions from a high to a low capacitance. Such extreme suddenness might not be expected when emission velocities are taken into account. Moreover, few valves saturate over their entire cathode area at one value of V_S. We should, therefore, not expect such sudden change near saturation, although fairly rapid changes of surprising magnitude could be expected.

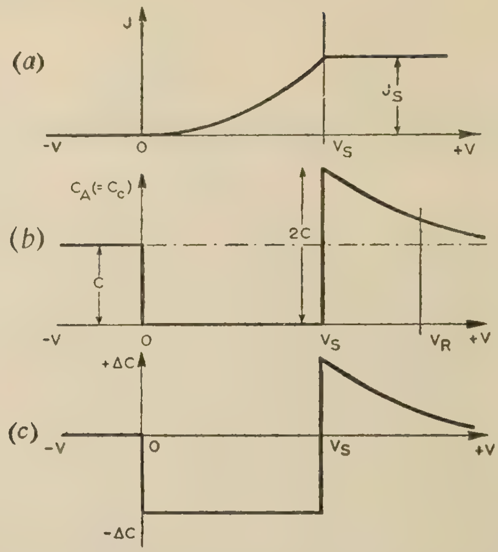


Fig. 2.—Variation of three parameters with anode voltage.

(a) Anode current density.
(b) Capacitance.
(c) Change in capacitance.

(6) DISCUSSION

The conductance, g, of a diode is so great under space-charge-limited conditions, that the time-constant C/g is of the order of the time of flight of the electrons. The effects discussed here could therefore be examined experimentally only in a range of frequencies one or two orders lower than the reciprocal of the time of flight. In such ranges the vanishing of C_A might have some utility, especially in the production of a purely resistive load for use in terminating transmission lines.

The phase changes occurring when the frequency is approximately equal to the time of flight would cause the capacitive currents discussed here to become negative-resistive currents. Oscillations are, in fact, maintained readily under near-space-charge-limited conditions, as in beam tetrodes operating near the knee voltage.[2]

The action of the voltage at the gaps of the resonators of klystrons can be studied in the light of the results given here; the matter is, however, complicated and has been worked out only in a qualitative way found to give general agreement with experience.

We should notice first that at any particular instant the beam current, whether or not it is modulated, is a saturated current not subject to variation on account of the gap voltage, except in one extreme condition to be discussed later. The gap faces can therefore be regarded as a planar diode, the electrons having a large constant emission energy equal to the direct resonator potential, V_R, say.

It is convenient to consider the action of space charge during the accelerating and the retarding portions of the gap-voltage cycle, taking the accelerating portion first. At any particular instant, for which the gap voltage is V, an increase in potential δV causes an increase in the electron speed. This results in an excess flow of electrons to the more positive electrode. The capacitance is therefore as for a saturated diode, greater than the cold capacitance by an extent which depends on the beam current, the gap-width and the electron speed. If the gap voltage, V, is always much less than V_R, the gap capacitance is very nearly a constant. We can thus conclude that the gap voltage will be sinusoidal, so that

$$V = V_0 \sin \omega t$$

This condition can be regarded as being represented approximately in Fig. 2 by a taking V_R at some position so that $V_R > V_S$, and V as being a small excursion about V_R, so that the gap capacitance does not vary appreciably during a cycle.

If, however, V is not much less than V_R, then during a positive excursion of V the gap capacitance, while remaining greater than the cold capacitance, is decreasing as V increases, as is also illustrated roughly in Fig. 2. The voltage excursion, V, can now no longer be sinusoidal but must be somewhat more peaky, the maximum energy of the magnetic field building up a greater voltage across the gap than the value which would have pertained had the gap capacitance remained constant.

The inverse effect occurs on negative excursions of V. The gap capacitance increases as the space charge increases as the electrons are slowed down. Should a virtual cathode be just not formed before the exit face of the resonator, Fig. 2 shows that the capacitance when $V = -V_R$ will be just equal to twice the cold capacitance. The increase in capacitance on the negative excursion of V is greater than the decrease on the positive excursion, the curve shown in Fig. 2 being steeper as the electron speed decreases.

As a consequence of the increase in capacitance, the energy of the magnetic field will be incapable of developing a voltage as high as V_0 across the gap. In an extreme case it may be reduced to one-half when the virtual cathode is just not formed.

The above considerations apply even when the beam is not modulated in density. The effects are not negligible in amount, since in all klystrons the tunnel is designed so as to be as fully as possible space-charge limited, and space charge in the gaps cannot therefore be entirely neglected. The situation is, in fact, emphasized when the beam is modulated and power is being drawn from the catcher. Then it is commonly the case that the drift length and modulation voltage are so chosen that during about nine-tenths of a cycle the beam current is about half its average value, and during the remaining tenth, centred on the maximum retarding voltage at the catcher gap, is about ten times the average value.

The peak of the positive excursion under these conditions would then exceed the sinusoidal voltage peak, $+V_0$, by a moderate amount. However, the peak of the retarding voltage may be very considerably less than $-V_0$, not only because the capacitance increases more rapidly, but also because the current density itself is very high at the time the peak voltage occurs.

Now in Webster's theory of the klystron[3] it is assumed that the gap voltage is sinusoidal; this will have two effects, namely

(*a*) It will ignore the slightly greater power carried off on the beam on account of the positive peak voltage being slightly greater than $+V_0$.

(*b*) It will not allow for the fact that the power given up by the beam during the most effective part of the cycle is considerably less than that calculated for a sinusoidal gap voltage. It is this power which is the most important component of the output of the klystron.

We can therefore expect that the efficiency calculated from Webster's theory will not be attained, unless the space charge in the gap is made very small. This indicates that smaller gap spacings must be used than would ordinarily be used to get sufficiently good coupling of the beam to the resonator.

Another effect also arises when extreme conditions are emploЯd. If the drift length is sufficiently great that a small velocity modulation will cause an intense bunch to form, then, when this reaches the gap of a suitably matched resonator, the voltage may fall to zero and a virtual cathode may form in front of the further face of the gap. The capacitance will then fall rapidly to zero, as shown in the discussion of the diode, and the voltage on it will consequently start to go negative very rapidly. The entire beam will then be rejected in a time comparable with the time of transit of electrons in the gap, i.e. commonly about 0·05 of a cycle. The capacitance will then revert quickly to its cold value, and a very rapid series of practically aperiodic transients will develop until the most negative point of the cycle is passed. Such instability will depend on the use of a long drift space, since it will not arise so readily when there is an appreciable range of velocities in the electron stream, which would be necessary to produce as efficient a bunch of electrons in a short drift length.

The author once observed such instability which was at the time quite inexplicable in a low-voltage klystron, and it has also been reported in recent work on the development of a high-power klystron.[4] It would be reduced by a reduction in the catcher gap-length or by an artificial increase in its capacitance. These are not necessarily acceptable solutions, since both methods increase the losses in the resonators and the increase in the power absorbed from the beam may be lost in the resonator.

Space charge has a further effect on capacitance. Hitherto we have always considered the fluctuations of an electron current to be due to fluctuations in the rate of arrival of electrons at the anode. We now, however, have a situation in which there will be changes in capacitance due to electronic processes in the valve, and also therefore subject to fluctuations. It is therefore probable that the fluctuations under saturated conditions will decrease to the shot-noise level only when the anode voltage is very high compared with the saturation voltage. Furthermore, for diodes the fluctuation will be considerably greater than that calculated from fluctuation theory[5] for space-charge-limited conditions, since the fluctuations in capacitance will not necessarily be correlated with the fluctuations in current.

Finally, we shall indicate the application of the methods used here to electron-inertia effects.

It will be noticed that in developing expressions for C_A and C_C we have at no point introduced the cold capacitance, C, and subsequently added to this correction factors. The cold capacitance has come out of the equations as a limiting case for zero space charge, because we have taken into account the entire field between the electrodes, that due to space charges between the electrodes and the charges bound by them.

In dealing with the energetic problems due to the flow of electrons we must follow the same procedure, putting for the current

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

which is one of Maxwell's equations.

For $\boldsymbol{J}$ we can use

$$\boldsymbol{J} = \rho \boldsymbol{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and for $\boldsymbol{D}$ we must recognize that the displacement at any point is due to space charges and bound charges, putting

$$\boldsymbol{D} = \boldsymbol{D}_s + \boldsymbol{D}_b \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Then we get

$$\nabla \times \boldsymbol{H} = \rho \boldsymbol{v} + \frac{\partial \boldsymbol{D}_s}{\partial t} + \frac{\partial \boldsymbol{D}_b}{\partial t} \quad . \quad . \quad . \quad (16)$$

Taking the divergence

$$\begin{aligned} \nabla(\nabla \times \boldsymbol{H}) &\equiv 0 \\ &= \nabla(\rho \boldsymbol{v}) + \frac{\partial}{\partial t}\nabla \boldsymbol{D}_s + \frac{\partial}{\partial t}\nabla \boldsymbol{D}_b \quad . \quad . \quad (17) \end{aligned}$$

But

$$\nabla \boldsymbol{D}_s = \rho \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Therefore

$$\nabla(\nabla \times \boldsymbol{H}) \equiv 0 = \nabla(\rho \boldsymbol{v}) + \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial t}\nabla \boldsymbol{D}_b \quad . \quad . \quad (19)$$

From this expression we can remove the portion

$$\nabla(\rho \boldsymbol{v}) + \frac{\partial \rho}{\partial t} = 0 \quad . \quad . \quad . \quad . \quad . \quad (20)$$

which is the ordinary equation of continuity for fluid flow, so that we get,

$$\frac{\partial}{\partial t} \nabla \boldsymbol{D}_b = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

as might be expected, since $\nabla \boldsymbol{D}_b = 0$ in the space between the electrodes.

This indicates that the equation to be solved to study electron-inertia effects is the Maxwell equation given in the form making proper allowance for bound and free charges, namely eqn. (16).

One difficulty with the solution of this equation is that, while $\boldsymbol{D}$ may be calculated from

$$\boldsymbol{D} = -\epsilon_0 \operatorname{grad} V \quad . \quad . \quad . \quad . \quad . \quad (22)$$

the velocity is not uniquely determined by

$$eV = \tfrac{1}{2} m |\boldsymbol{v}|^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

This point has been mentioned previously, and it leads to the necessity to discuss only simple cases in which emission velocities and electrons flowing at different velocities in the same region are not taken into account.

In his monograph on electron-inertia effects, Llewellyn[6] uses the expression (13) to derive

$$\nabla \times \boldsymbol{H} = \rho \boldsymbol{v} + \frac{\partial \boldsymbol{D}}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

and puts $\rho = \nabla \boldsymbol{D}$, getting, for an electron flow in one direction and with single electron speed, a total current

$$\boldsymbol{I} = (\nabla \times H) = \frac{\partial D}{\partial x}\frac{dx}{dt} + \frac{\partial D}{\partial t} = \frac{dD}{dt} \quad . \quad . \quad (25)$$

In this, $\boldsymbol{D}$ is essentially restricted to the value which we have called $\boldsymbol{D}_s$, and so eqn. (25) is incomplete. Consequently the well-known equation resulting from eqn. (25),

$$\boldsymbol{I} = k\dddot{x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

expressing the current as the third derivative of the electron position, is generally in error. It is valid only if there are no electrodes on which charges can be bound by the space charge. This is so severe a restriction as to prevent the application of the results to any actual device. In consequence of this we must reject Llewellyn's conclusions that the capacitance of a space-charge-limited diode at low frequencies is $\frac{3}{5}C$ and that no marked change in the capacitance takes place as we move from the space-charge-limited to the saturated régime.

(7) REFERENCES

(1) van der Ziel: 'Extension and Application of Langmuir's Calculation on a Plane Diode with Maxwellian Velocity Distribution of the Electrons', *Philips Research Reports*, 1945–46, **1**, p. 97.

(2) Bull, C. S.: 'Space-Charge Effects in Beam Tetrodes and other Valves', *Journal I.E.E.*, 1948, **95**, Part III, p. 17.

(3) Webster, D. L.: 'Cathode-Ray Bunching', *Journal of Applied Physics*, 1939, **10**, p. 50.

(4) Lawson, J. D., *et al.*: 'The Design and Performance of a Demountable High-Powered X-Band Klystron', *Journal of Electronics*, 1955, **3**, p. 333.

(5) Bull, C. S.: 'The Noise of Valves', *British Journal of Applied Physics*, 1954, **5**, p. 127.

(6) Llewellyn, F. B.: 'Electron Inertia Effects' (University Press, Cambridge, 1941).

(7) Ramo, S., and Whinnery, J. R.: 'Fields and Waves in Modern Radio' (Wiley, New York, 1944).

621.317.75

Paper No. 2306 M
Feb. 1957

THE ULTIMATE PERFORMANCE OF THE SINGLE-TRACE HIGH-SPEED OSCILLOGRAPH

By M. E. HAINE, M.Sc., F.Inst.P., Member, and M. W. JERVIS, M.Sc.Tech., Associate Member.

(The paper was first received 4th July, and in revised form 3rd December, 1956. It was published in February, 1957, and was read before a joint meeting of the MEASUREMENT AND CONTROL SECTION *and the* RADIO AND TELECOMMUNICATION SECTION *12th February, 1957.)*

SUMMARY

The limitations to writing speed and deflection sensitivity of the single-trace high-speed oscillograph are analysed in terms of theoretical electron-gun performance and the spherical aberration of practical electron lenses. It is shown that considerable improvements over the best present-day designs are theoretically possible. The results indicate the design trends which should achieve the improved performance.

LIST OF PRINCIPAL SYMBOLS

τ = Temporal resolution, sec/spot-width.
n_s = Deflection sensitivity, spot-widths/volt.
σ = Charge density, coulombs/cm^2.
J = Current density in beam, amp/cm^2.
α = Semi-angle of beam, rad.
J_c = Cathode current density, amp/cm^2.
T = Cathode temperature, deg K.
V_0 = Electron-gun accelerating voltage, volts.
k = Boltzmann's constant (1/11 600 eV/deg K).
β = Gun brightness, amp/cm^2/steradian.
d = Spot diameter, cm.
c_s = Spherical aberration constant, cm.
v = Distance between lens and image, cm.
u = Distance between lens and object, cm.
f = Focal length of lens, cm.
s = Lens pole-piece separation, cm.
D = Lens bore diameter, cm.
t = Electrostatic-lens electrode thickness, cm.
z_p = Deflector-plate length, cm.
z_s = Distance between deflector plate and screen, cm.
y_p = Deflector-plate separation, cm.
e = Electronic charge, coulombs.
d_m = Minimum beam diameter, cm.
l = Converging distance of electrons, cm.
I_b = Beam current, amp.
σ_d = Charge density, direct recording, for density of 0·2, coulombs/cm^2.
σ_i = Charge density, indirect recording, for density of 0·2, coulombs/cm^2.
V_1 = Post-deflection accelerating potential, volts.
σ' = Charge density with image intensifier, coulombs/cm^2.

(1) INTRODUCTION

There have been many papers describing high-speed single-trace oscillographs, and excellent reviews are given by Craggs and Meek[1] and others.[2, 3] Before 1940, continuously-evacuated oscillographs were normally used for recording high-speed transients, but subsequently, sealed-off tubes were shown to be capable of a performance sufficient for most, if not all, practical purposes. The relative cheapness and convenience of the sealed tube has obviously encouraged this view. Surprisingly, there has been little apparent effort to analyse either the relative theoretical performance of the two methods, or the limits of performance of either. The paper attempts to analyse the limitations of the two methods to show how far future improvements might go. It deals mainly with limitations due to the beam and recording media, rather than to the deflector plates and their characteristics, since the latter have been dealt with fairly fully,[2] but some comments are made on the significance of certain aspects of the deflector-plate design.

The definitions of writing speed and sensitivity used will be as follows. Linear writing speed is replaced by 'temporal resolution' or just 'resolution'—the minimum time taken for the spot to move one diameter and still be recordable. The deflection sensitivity is taken as the number of spot widths of deflection for one volt applied to the deflector plates.

With modern oscillographs, temporal resolutions in the region of 10^{-12} sec can be obtained with deflection sensitivities in the range 0·03–1 spot-widths/volt. This resolution will give single-stroke oscillograms of a 10 000 Mc/s wave of peak-to-peak amplitude 30 spot widths, corresponding to about 1 kV. Such a resolution would be capable of displaying a pulse with a rise time of 10^{-10} sec and a height of 100 spot widths. These are values of the order of those which occur in investigations of the build-up of microwave magnetron waveforms and spark discharges. Improved resolutions would enable higher-quality oscillograms to be obtained and faster rise-times recorded.

(2) TEMPORAL RESOLUTION

The temporal resolution of the oscillograph depends entirely on the density of electrons which must fall on the photographic plate in the direct-recording demountable oscillograph or on the phosphor in the sealed-off oscillograph in order to produce adequate blackening of the photographic plate (in the latter case after the fluorescent light has been focused by an optical lens). This quantity can of course be measured and will depend upon such factors as the electron energy, the photographic-plate sensitivity and the efficiency of the optical system where used.

If the spot at the phosphor or photographic plate is formed by a beam of current density J, the charge density falling on any point in the trace is given by

$$\sigma = J\tau \qquad (1)$$

This expression assumes a spot of uniform current density, but, in practice, this distribution will more usually have a Gaussian form, which modifies eqn. (1) by a numerical factor which is, however, sufficiently close to unity to be neglected if the spot width is defined as the width at half height, and J is measured at the centre of the spot.

The maximum current density obtainable from an electron gun depends only on the maximum usable beam angle, the cathode current density and temperature of the cathode and the accelerating voltage according to the relation[4]

$$J_{max} = \frac{J_c V_0}{\pi k T}\pi\alpha^2 \qquad (2)$$

Mr. Haine is, and Mr. Jervis was formerly, in the A.E.I. Research Laboratory.
Mr. Jervis is now with the A.E.I.–John Thompson Nuclear Energy Co., Ltd.

The factor $J_c V_0/\pi kT = \beta$ gives the current density per unit solid angle or brightness of the system. This quantity is invariant throughout the system in the absence of beam-limiting apertures.

Putting $J = \beta\pi\alpha^2$ in eqn. (1) gives

$$\sigma = \beta\pi\alpha^2\tau$$

or

$$\alpha = \left(\frac{\sigma}{\pi\beta\tau}\right)^{1/2} \text{ and } \tau = \sigma/\pi\beta\alpha^2 \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This equation defines the beam semi-angle necessary to give a temporal resolution of τ when the charge density required to give a recordable image is σ and the electron-gun brightness is β. It will be noted that the temporal resolution is independent of the spot size.

If the conditions are adjusted to give maximum brightness and a minimum required charge density, the resolution can be improved only by increasing α, the maximum value of which is usually limited either by the spherical aberration of the imaging lens, or by deflection defocusing errors. The latter are not usually significant in an oscillograph of the kind under consideration, since wide-angle deflection is unnecessary.

(3) SPHERICAL ABERRATION

The maximum beam intensity as limited by spherical aberration will be reached when the disc of minimum confusion due to the spherical aberration becomes about equal in size to the spot. The diameter of this disc is given by $\frac{1}{2}c_s\alpha^3(v/f)^4$, so that the maximum allowable beam angle is given by

$$\alpha_m = [(2d/c_s)(f/v)^4]^{1/3} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Substituting this value in eqn. (3) gives:

$$\tau = (\sigma/\pi\beta) \times (c_s/2d)^{2/3} \times (v/f)^{8/3} \quad . \quad . \quad . \quad . \quad (5)$$

The resolution time now becomes dependent on the spot size as well as on v/f.

From this expression the limits of resolution for a practically feasible oscillograph can be deduced. Consider a system such as that shown in Fig. 1. S is here the electron source (which

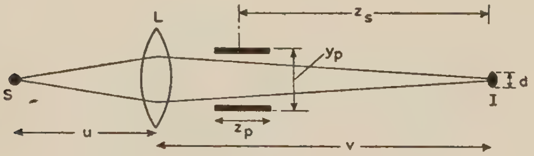


Fig. 1.—Basic oscillograph electron-optical system.

may be virtual or may be produced by a previous lens or lens system), L is the main focusing lens of focal length f and spherical aberration constant c_s, and I is the image: SL $= u$ and LI $= v$. The final spot diameter is d and it is assumed the spot diameter at S can be adjusted to give any desired value of d.

The minimum value of τ is obtained with the lens of shortest focal length, and hence minimum spherical aberration, with a maximum value of d and minimum magnification. The substitution of limiting values shows that a value of τ below 10^{-17} sec is possible, but the arrangement would then be of little practical use as an oscillograph, since d would be large and v small, and hence the deflection sensitivity would be very low. Before attempting to evaluate τ, therefore, some consideration must be given to the question of deflection sensitivity, which has an importance in oscillography second only to resolution.

In any practically useful system the focusing lens is a weak one (i.e. the focal length is long compared with the axial extent

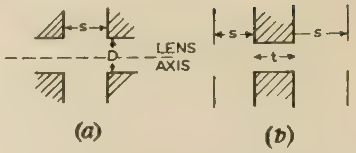


Fig. 2.—Electron lenses.
(a) Magnetic.
(b) Electrostatic.

of the field). The spherical aberration of weak electron lenses can be expressed by simple analytic expressions. For a magnetic lens with pole pieces of bore diameter D and spacing s [Fig. 2(a)] the expression is[5]

$$c_s/f = 5[f/(s + D)]^2 \quad . \quad . \quad . \quad . \quad (6a)$$

For the electrostatic einzel lens a similar expression holds.[6] If the gap spacing is S and the central electrode thickness is t [Fig. 2(b)],

$$c_s/f = 3 \cdot 5[f/(S + \tfrac{1}{2}t)]^2 \quad . \quad . \quad . \quad . \quad (6b)$$

In the following analysis the former expression will be used, though it is seen that the results for electrostatic lenses will be little different. The focal length of the weak magnetic lens is given by

$$f/(s + D) = 25V_0/(NI)^2 \quad . \quad . \quad . \quad . \quad (6c)$$

where NI is the excitation m.m.f. applied to the magnetic gap.

Substituting eqn. (6a) into eqn. (5) gives

$$\tau = (\sigma/\pi\beta)\left[\frac{2 \cdot 5}{d}(f/s + D)^2\right]^{2/3} v^{8/3}/f^2 \quad . \quad . \quad (7)$$

Assuming for the moment that the ratio $f/(s + D)$ is fixed, it is seen that a maximum value of f gives minimum resolution time. This means that f should be as nearly equal to v as possible, and therefore u should be very large. Practically, it is undesirable to make u too large, and little is gained by making it greater than four or five times v.

Putting $u = 4v$ gives $f = 0 \cdot 8v$, and eqn. (7) becomes

$$\tau = 0 \cdot 7\frac{\sigma}{\beta}\left(\frac{v}{s + D}\right)^{4/3}\left(\frac{v}{d}\right)^{2/3} \quad . \quad . \quad . \quad . \quad (8)$$

(4) DEFLECTION SENSITIVITY

A deflector system (see Fig. 1) with a pair of plates of length z_p and spacing y_p placed in close proximity to the lens (immediately before or after) will have a deflection sensitivity given approximately by

$$n_s = z_p z_s/2y_p V_0 d \text{ spot widths per volt} \quad . \quad . \quad . \quad (9)$$

It is clear that the dimensions of the deflector system must be chosen with regard to the space available, limited axially by the dimension v and radially (spacing) by the beam diameter. In addition, circuital considerations must come in. The plate capacitance and inductance and the transit-time effects become very important as the frequency of the deflecting voltage increases. For the present, however, the latter effects will be disregarded, except in so far as overall size considerations come in.

The plate dimensions can be expressed in terms of v and $z_s\alpha$ (the beam diameter at the plates); thus, we can introduce three numerical constants P, Q and R such that

$$z_p = Pv \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$z_s = Qv \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and

$$y_p = Rz_s\alpha$$
$$= RQv\alpha \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

After substitution of eqn. (6a) in eqn. (4), and putting $f = 0{\cdot}8v$, α_m is found to have a value given by

$$\alpha_m = \frac{0{\cdot}68(d/v)^{1/3}}{[v/(s+D)]^{2/3}} \quad . \quad . \quad . \quad . \quad (13)$$

Hence

$$y_p = \frac{0{\cdot}68RQv(d/v)^{1/3}}{[v/(s+D)]^{2/3}} \quad . \quad . \quad . \quad . \quad (14)$$

Substituting these values into eqn. (9) gives

$$n_s = 0{\cdot}73(P/R) \times (v/d)^{4/3}[v/(s+D)]^{2/3}/V_0 \quad . \quad (15)$$

Eliminating v/d between eqns. (8) and (15) gives

$$\tau = 0{\cdot}8(\sigma/\beta) \times (n_s V_0)^{1/2}[v/(s+D)]/(P/R)^{1/2} \quad . \quad (16)$$

So far, the expressions developed set no absolute value to the size of the oscillograph. Scaling all dimensions in proportion would leave the performance, as so far defined, unchanged. The size is, however, very largely affected by factors limiting the usable spot size. In the first place, the spot size must clearly not be less than the resolving power of the photographic plate, phosphor, or other receiving screen used. Secondly, the spot size is limited by space-charge spreading of the beam, as is shown in the next Section.

(5) SPACE-CHARGE LIMITATION

A convergent beam of electrons can never contract below a certain limiting diameter because of mutual repulsion of the electrons. The minimum diameter which can be obtained with a beam of energy eV_0 and semi-angle α, converging over a distance l is given by[7]

$$d_m = 2l\alpha \exp - (3{\cdot}24 \times 10^{-5}V_0^{3/2}\alpha^2/I_b)$$

It will be noted that d_m is critically dependent upon I_b; above a certain current, the minimum diameter of a beam, otherwise limited only by geometric optical considerations, will suddenly increase in diameter by an amount increasing rapidly with the current. Thus, for a beam of given angle and voltage there is a limiting current beyond which the beam may be said to explode or become unstable. This current is given from the above expression by

$$I_b < 1{\cdot}4 \times 10^{-5}V_0^{3/2}\alpha^2/\log_{10}\frac{2l\alpha}{d_m} \quad . \quad . \quad . \quad (17)$$

Since the value inside the logarithm has only a small effect upon the value of the logarithm, very approximate values of l, α and d_m can be inserted, e.g. $l = 20$ cm, $\alpha = 0{\cdot}02$ rad and $d_m = 2 \times 10^{-3}$ cm. The logarithm then has the value $2{\cdot}6$, and we can write

$$I_b < 5 \times 10^{-6}V_0^{3/2}\alpha^2$$

but since

$$\beta = 4I_b/\pi^2\alpha^2d^2$$

$$\beta < 2 \times 10^{-6}V_0^{3/2}/d^2$$

or

$$d < 1{\cdot}4 \times 10^{-3}V_0^{3/4}/\beta^{1/2} \quad . \quad . \quad . \quad . \quad (18a)$$

putting

$$\beta = \frac{J_cV_0}{\pi kT} \text{ with } k = \frac{1}{11\,600}\text{ eV/deg K}$$

$$d < 2{\cdot}3 \times 10^{-5}V_0^{1/4}T^{1/2}/J_c^{1/2} \quad . \quad . \quad . \quad (18b)$$

Since the potential comes only in the power of 1/4, its value is of little importance in this expression. It can therefore be given a somewhat arbitrary value of 10 kV. Table 1 then gives the maximum allowable spot size in microns as a function of current density for $T = 1\,000°$ and $3\,000°$ K.

The resolution of a photographic plate to electrons is about 20 μ and that of a phosphor is about 80 μ.

Table 1

ALLOWABLE SPOT SIZE

J_c	Spot size	
	at 1 000° K	at 3 000° K
amp/cm²	μ	μ
10	23	40
1	73	125
0·1	230	400

(6) OVERALL SIZE OF OSCILLOGRAPH

The size of the oscillograph can now be derived in terms of the dimension v from eqn. (15).

$$v = \frac{1{\cdot}2(V_0n_sR/P)^{3/4}d}{[v/(s+D)]^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

For minimum size, d must take the minimum value determined by the plate or screen resolution, but it must not exceed the value given in Table 1. For a phosphor at 50 kV it is about[8] 80 μ, but this figure may well be nearer 20 μ at 10 kV for a carefully prepared screen. Since the size of the oscillograph is proportional to the spot size, an advantage in size may be gained, when high-voltage operation is required, by the use of direct recording.

(7) THE RECORDING MEDIUM

It is of interest to compare the sensitivity of the direct to the indirect method of recording.

(7.1) Direct Photography

The sensitivity of the photographic emulsion to electron irradiation has been studied by a number of workers. It is important to note in the first place that the wide range of speeds found with different emulsions exposed to light no longer holds for electron irradiation. Digby, Firth and Hercock[9] have examined a representative series of emulsions, and, in common with others, found that for normal types of emulsion, a range of only about 10 : 1 exists between the fastest and slowest emulsions. The purpose under discussion requires a fast plate, and Fig. 3 shows the charge density required to give a density of $0{\cdot}2$ above fog as a function of electron energy for the fastest plate for electrons. Between 20 and 50 kV, σ is given approximately by

$$\sigma_d = 1{\cdot}2 \times 10^{-7}/V_0 \text{ coulomb/cm}^2 \quad . \quad . \quad . \quad (20)$$

(7.2) Use of Phosphor (Indirect Photography)

For indirect photography the phosphor efficiency, the camera aperture and film speed for light affect the final sensitivity. The efficiency of the phosphor in terms of light output per unit electron energy of irradiation can be as high as 30% for a thick screen viewed from the side irradiated. The efficiency is independent of electron energy between 10 and 60 kV, and falls above 60 kV. Some loss results for a phosphor viewed from the front, but, if an optimum thickness is chosen and a reflecting layer is put on the back, the loss should be small. The minimum energy detectable on a very fast photographic plate (38° Scheiner) is about 4×10^{-7} lumen-sec/cm². On this basis, the value of σ for the indirect method is given by

$$\sigma_i = (2{\cdot}1 \times 10^{-9}/V_0)[4F^2(M+1)^2 + 1] \text{ coulomb/cm}^2 \quad . \quad (21)$$

where F is the F-number of the optical lens used, and M is the magnification between the phosphor and the photographic

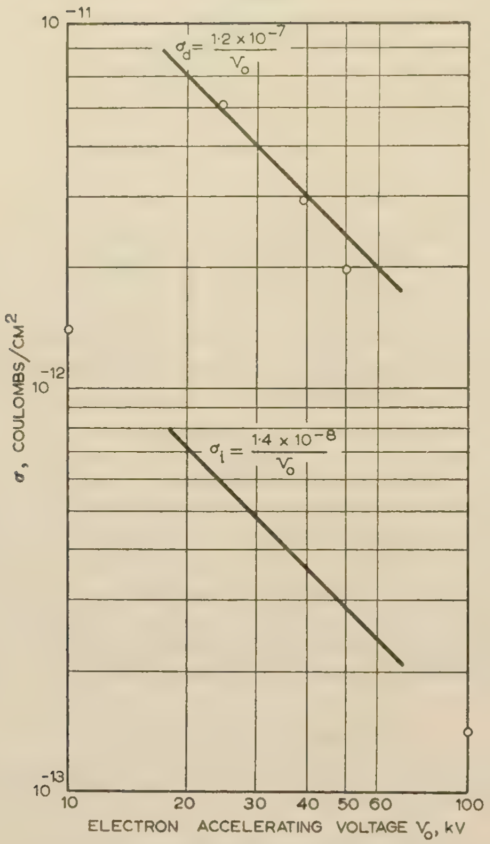


Fig. 3.—Variation of recording sensitivity by direct and indirect photography with electron accelerating voltage.

—○—○— Experimental points[9] for a density of 0·2.

film. For an $F.1$ lens operating with a 4 : 1 demagnification and neglecting lens transmission losses, $4F^2(M + 1)^2 + 1 = 7{\cdot}25$,

and $\qquad \sigma_i = 1{\cdot}5 \times 10^{-8}/V_0$ coulomb/cm² . . . (22)

Thus, the indirect method is capable of about ten times the sensitivity of the direct method. This is, however, true only if the 4 : 1 optical demagnification is possible: this is so only if the spot size used is at least four times the photographic plate resolution—a condition not realized for spots of 20–50μ and high-speed plates. Because of this, the light loss in the lens, some loss due to reciprocity failure and the loss of efficiency in a transmission phosphor, the indirect method will at best equal the direct.

In addition, screen saturation is not allowed for. The current density in the spot is given by σ/τ. Phosphors are normally expected to saturate at a few milliamperes per square centimetre at most. However, experiments by Einstein[10] have shown that, under short pulse conditions, saturation does not occur until the current density is in excess of at least 1 amp/cm². Hence the loss due to saturation should not be serious for resolution times in excess of about 10^{-13}–10^{-12} sec. At worst, it appears that the direct and indirect methods of recording are comparable in sensitivity.

(8) EVALUATION OF RESULTS

It is now possible to evaluate the limiting performance. First, some reasonable values for P and R must be fixed: z_s clearly must be small compared with v, and a value of $P = 0{\cdot}2$ seems perfectly reasonable; y_p must be greater than the beam diameter ($v\alpha$) and R might be given a value of 2, although if wide deflection angles are required, this could be too small. It is, however, not an unreasonable value to take in estimating the limiting performance. The value to be set for $v/(s + D)$ is limited on the one hand by the question of field penetration from the lens, and on the other by physical size. A value of $v/(s + D) = 3$ seems quite possible, although it might lead to a rather large lens where high deflection sensitivity coupled with short resolution time is required.

Now $\qquad \tau = 8(\sigma/\beta)(n_s V_0)^{1/2}$ . . . . (23)

Putting $\qquad \sigma = 1{\cdot}2 \times 10^{-7}/V_0$ (direct photography)

and $\qquad \beta = \dfrac{J_c V_0}{\pi k T}$

gives $\qquad \tau = 2{\cdot}4 \times 10^{-10} T n_s^{1/2} / J_c V_0^{3/2}$ . . . (24)

Table 2 shows how this limiting value of τ varies with V_0 for three values of deflection sensitivity, for a cathode temperature of 1000° C, a cathode current density of 1 amp/cm² and a spot of 20μ diameter.

Table 2

LIMITING VALUES OF τ

V_0	n_s	τ	α_n	I_b	J	v
kV	spot widths/volt	sec $\times 10^{-14}$	rad $\times 10^{-3}$	μA	amp/cm²	cm
10	0·1	8·2	31	461	147	1·4
10	1·0	26	17	145	46	7·8
10	10·0	82	10	46	14·7	44
30	0·1	1·6	24	780	250	3·1
30	1·0	5	13	250	80	18
30	10·0	16	8	78	25	102
50	0·1	0·75	21	1 000	320	4·6
50	1·0	2·36	18	320	102	26
50	10·0	7·5	7	100	32	153

It will be seen that, without making any experimentally impossible assumptions, the analysis predicts that greatly improved oscillographs should be possible, with recording speeds 100 times higher than the best values so far obtained and deflection sensitivities which compare favourably with those obtained on the fastest available oscillographs. For the highest speeds and highest deflection sensitivities shown in Table 2, the tube length is becoming rather excessive, requiring rather a large lens, but the dimensions are still well within practical limits, although for indirect recording ($d = 80\mu$), the size might be prohibitive; however a further factor of 10 in improvement in resolution time should be possible by increasing the current density from 1 to 10 amp/cm.²

(9) POST-DEFLECTION ACCELERATION

Increased deflection sensitivity can be obtained by the well-known process of post-deflection acceleration, in which the deflection takes place in a region where the beam is at a potential V_1, with respect to the cathode, whereas the beam is accelerated to a higher potential V_0 before the final image is reached. With post-deflection acceleration an improvement in deflection sensitivity by a factor $\sqrt{(V_0/V_1)}$ is obtained,[11] as compared with an oscillograph operating at V_0 in the deflection and recording regions, provided that the final beam angle and spot size are unchanged and that the optical system is free from aberration.

The application of the method usually results in a gain smaller than that predicted,[12, 13] but there seems no reason why this should be so in a correctly designed oscillograph. The main error that has been made in past designs is using too large a spot size and screen, so that a large deflection angle is required. It then becomes difficult to arrange a suitable post-deflection acceleration system.

(10) POSSIBLE APPLICATION OF THE IMAGE INTENSIFIER

Now that the image intensifier has reached a practically useful stage in its development, it is of interest to consider whether its application in cathode-ray oscillography may not lead to the possibility of even greater speeds and deflection sensitivity. An ultimate limitation will be reached even with an ideal intensifier system when the number of electrons per spot is so small that statistical fluctuations become objectionable. This limit would occur when such fluctuation exceeded about 10%, i.e. when the number of electrons per spot was less than 100.

Thus, a new limiting value of σ' can be defined from

$$\sigma' = \frac{100 \times 1{\cdot}6 \times 10^{-19}}{\pi d^2/4} \text{coulomb/cm}^2 \quad . \quad . \quad . \quad (25)$$

where $1{\cdot}6 \times 10^{-19}$ coulomb is the charge on the electron. It is seen that high sensitivity (low σ') requires large spot size. Since the spot size is, however, limited by space-charge spreading, the value of σ' is limited. With the limiting value of $d = 20\mu$, σ' becomes 5×10^{-12} coulomb/cm^2. This is already twice the value needed for direct recording on the fastest available emulsion at 50 kV. Hence, this plate is ideally sensitive, and already just about at the limit set by quantum noise.

To obtain a gain by the use of an intensifier, the spot size must be chosen so that $\sigma' < \sigma$. For a tenfold improvement over the special rapid plate, σ' must be 2×10^{-13} coulomb/cm^2 at 50 kV. This requires a spot $0{\cdot}1$ mm in diameter [eqn. (25)].

Table 1 shows that the cathode current density is then limited to 1 amp/cm^2. Thus, the gain is no more than would be obtained by increasing the current density from 1 to 10 amp/cm^2 with the fastest available emulsion, and a spot size of 20μ. On the other hand, an intensifying or image-transfer system has several practical advantages, even if it offers little promise of improving the oscillograph performance, except possibly if high deflection sensitivity is especially important and temporal resolution can be sacrificed in its favour. In this case, it may be possible to make use of the higher resolution which might be obtained from a pick-up tube, and so use a smaller spot size in the oscillograph.

If the intensifying system comprises a scanned target of the Vidicon type,[14] it is possible that storage could be achieved by the use of a screen material such as antimony trisulphide which, in some forms, has a very long persistence of induced conductivity.[15]

(11) PRACTICAL CONSIDERATIONS

(11.1) Electron Gun

In the foregoing analysis it has been assumed that an electron gun giving a theoretically best performance is practically attainable, and there is no reason to suppose that such is not the case. It has been demonstrated that theoretical performance can be obtained from a temperature-limited cathode gun,[16] and there is little doubt that equivalent performance is also possible from a space-charge-limited gun. This is particularly so if no negative grid is used, since most of the aberrations of the negative-grid gun as used in the cathode-ray tube are almost certainly due to the strong field curvature immediately in front of the cathode, resulting from the presence of the grid. A grid is not necessary in a single-trace oscillograph, where a beam trap can be used.

(11.2) Deflector Systems

For the very high speeds predicted in the foregoing analysis, the electron transit time in the deflector plates becomes of major importance, unless very short plates are used with the resultant loss of deflection sensitivity. The transit-time effect can be mitigated by using travelling-wave deflector systems.[2, 17] It is not intended to analyse the limitations of such deflector systems in detail, but the following discussion will give some indication of these.

In the simplest form of travelling-wave deflector system, the plates are divided in N sections (Fig. 4) joined by inductors to

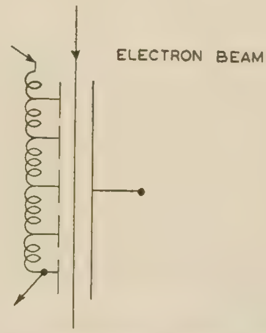


Fig. 4.—Basic travelling-wave deflector-plate system.

form an artificial transmission line. The inductances, L, are chosen to give a wave propagation velocity such that the time for it to travel between sections is equal to the corresponding time for the electrons to travel the same distance. The electrons thus 'see' the wave undistorted by transit-time effects. If the time for the electron to travel the full length of the deflector system is t_0, the above condition is met when

$$t_0 = N\sqrt{(LC)} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

where C is the plate-section capacitance.

The transmission of such a line falls off at frequencies above a cut-off value given by

$$f_c = 1/\pi\sqrt{(LC)} \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Combining this with eqn. (26) gives

$$f_c = N/\pi t_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

from which it is seen that higher frequencies require more sections. There is, however, a limit to the number of sections which can be used, because, as the sections become smaller, the capacitance between them becomes comparable with the section capacitance, resulting in serious inter-section coupling which upsets the function of the line. This effect can to a certain extent be reduced by suitable design, but a final limitation arises when the free-space wavelength corresponding to the applied frequency becomes comparable with the plate spacing.

Apart from this limitation, the use of a travelling-wave deflector system implies the absorption of power corresponding to that absorbed in a resistance equal to the characteristic impedance of the transmission line [$Z_0 = \sqrt{(L/C)}$], and it can readily be shown that this cannot practically be increased above a few hundred ohms. Thus, considerable power is required to drive the deflector plates. This is essentially a limitation to the transmission of high-frequency voltages, for such can be achieved only from a sufficiently low-impedance source. As frequency rises, the act of transmission of voltage is essentially coupled with the transmission of power. It will be noted that the power required to drive the oscillograph is not necessarily absorbed therein, and can, in general, be passed on for application elsewhere.

Thus, a distinct limitation is set to the application of the very high recording sensitivities shown possible. This limitation will be lessened as more effective travelling-wave amplifiers and broad-band travelling-wave tubes become available.

The limitation in no way lessens the value of the analysis, which, apart from showing the ultimate limitations, shows that, for much lower recording sensitivities, considerable economy in size, etc., could theoretically be effected.

(12) CONCLUSIONS

The limiting performance of a single-trace high-speed oscillograph is derived from first principles. The recording speed or resolution is shown to be limited by the gun brightness, which has a theoretical limit, or by space-charge spreading, and by the spherical aberration of the imaging lens and the recording sensitivity. Space-charge spreading can be made negligible for a given electron-gun brightness by the use of a sufficiently small spot. Conditions can be chosen for a minimum effect of spherical aberration. Deflection sensitivity depends largely on spot size, as suggested by von Ardenne[18, 19] and Lee,[20] the ultimate limit of spot size being set by the resolution of the recording screen. Recording sensitivities are already near the fundamental limit set by quantum noise.

Improvements of 10–100 times the present time resolution is predicted as a possibility without loss of deflection sensitivity.

(13) ACKNOWLEDGMENTS

Acknowledgment is made to Dr. T. E. Allibone, Director of the A.E.I. Research Laboratory, for permission to publish the paper.

(14) REFERENCES

(1) Craggs, J. D., and Meek, J. M.: 'High Voltage Laboratory Technique' (Butterworth, London, 1954).

(2) Lewis, I. A. D., and Wells, F. H.: 'Millimicrosecond Pulse Techniques' (Pergamon Press, London, 1954).

(3) Hardy, D. R., Jackson, B., and Feinberg, R.: 'The Recording of High-Speed Single-Stroke Electrical Transients', *Electronic Engineering*, 1956, **28**, pp. 8 and 75.

(4) Langmuir, D. B.: 'Theoretical Limitations of Cathode-Ray Tubes', *Proceedings of the Institute of Radio Engineers*, 1937, **25**, p. 977.

(5) Liebmann, G., and Grad, M. E.: 'Imaging Properties of a Series of Magnetic Electron Lenses', *Proceedings of the Physical Society*, 1951, **64B**, p. 956.

(6) Archard, G. D.: 'Focal Properties and Spherical Aberrations of the Three-Electrode Electron Lens', *British Journal of Applied Physics*, 1956, **7**, p. 330.

(7) Thompson, B. J., and Headrick, L. B.: 'Space Charge Limitations on the Focus of Electron Beams', *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 318.

(8) von Borries, B.: 'Die Energetischen Daten und Grenzen der Übermikroskopie', *Optik*, 1948, **3**, p. 321.

(9) Digby, N., Firth, K., and Hercock, R. J.: 'The Photographic Effect of Medium Energy Electrons', *Journal of Photographic Science*, 1953, **1**, p. 194.

(10) Einstein, P. A.: 'The Light Output of Some Phosphors excited with Electrons of High Current Density', *British Journal of Applied Physics*, 1957, **8**, p. 190.

(11) Pierce, J. R.: 'After Acceleration and Deflection', *Proceedings of the Institute of Radio Engineers*, 1941, **29**, p. 28.

(12) White, W. G.: 'Cathode-Ray Tubes with Post-Deflection Acceleration', *Electronic Engineering*, 1949, **21**, p. 75.

(13) Lempert, I. E., and Feldt, R.: 'The 5RP Multiband Tube. An Intensifier Type Cathode-Ray Tube for High-Voltage Operation', *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 432.

(14) Weimer, P. K., Forgue, S. V., and Goodrich, R. R.: 'The Vidicon Photoconductive Camera Tube', *RCA Review*, 1951, **12**, p. 306.

(15) Ansbacher, F., and Ehrenberg, W.: 'Electron Bombardment Conductivity of Dielectric Films', *Proceedings of the Physical Society*, 1951, **64A**, p. 362.

(16) Haine, M. E., and Einstein, P. A.: 'Characteristics of the Hot Cathode Electron Microscope Gun', *British Journal of Applied Physics*, 1952, **3**, p. 40.

(17) Honey, R. C.: 'A Travelling Wave Deflection System', *Transactions of the Institute of Radio Engineers*, MTT2, 1954, **2**, p. 2.

(18) Von Ardenne, M.: 'Der Elektronen-Mikrooszillograph', *Hochfrequentztechnik und Elektroakustik*, 1939, **54/6**, p. 181.

(19) Von Ardenne, M.: 'Ein Präzisions-Elektronenstrahloszillograph mit Wenigen Schreibfleckdurchmesser', *Nachrichtentechnik*, 1955, **5**, p. 481.

(20) Lee, G. M.: 'A Three-Beam Oscillograph for Recording at Frequencies of 10 000 Mc/s', *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 121.

[The discussion on the above paper will be found on page 390.]

621.317.755

Paper No. 2307 M
Feb. 1957

THE DESIGN AND PERFORMANCE OF A NEW EXPERIMENTAL SINGLE-TRANSIENT OSCILLOGRAPH WITH VERY HIGH WRITING SPEED

By M. E. HAINE, M.Sc., F.Inst.P., Member, and M. W. JERVIS, M.Sc.Tech., Associate Member.

(*The paper was first received* 4*th July, and in revised form* 3*rd December*, 1956. *It was published in February*, 1957, *and was read before a joint meeting of the* MEASUREMENT AND CONTROL SECTION *and the* RADIO AND TELECOMMUNICATION SECTION 12*th February*, 1957.)

SUMMARY

The paper describes the design and construction of a new high-speed cathode-ray oscillograph for recording single transients. A demountable construction is used which permits various types of deflector systems to be employed, and allows the current in the final spot to be measured by a Faraday cage moved to intercept the beam. Recording is by the direct action of electrons on a photographic plate.

The temporal resolution is measured from oscillograms of voltages of frequencies up to 9080 Mc/s, the limiting resolution being 2×10^{-13} sec per spot width: this figure is in agreement with design predictions.

LIST OF PRINCIPAL SYMBOLS

V_0 = Electron-gun accelerating voltage, volts.
J_c = Cathode current density, amp/cm^2.
β = Gun brightness, amp/cm^2/steradian.
τ = Temporal resolution, sec/spot-width.
σ = Charge density, coulombs/cm^2.
α = Beam semi-angle, rad.
N = Number of turns on lens.
I_l = Lens current, amp.
s = Lens pole-piece separation, cm.
D = Lens bore diameter, cm.
f = Lens focal length, cm.
f_t = Frequency of timing wave, c/s.
a = Peak-to-peak amplitude of trace, spot widths.
n = Number of peaks in length b of oscillogram.
dV/dt = Time-base speed, volts/sec.
I_d = Discharge current, amp.
C = Total time-base capacitance, farads.
d = Spot diameter, cm.
J = Current density in beam, amp/cm^2.
I_c = Collected current, amp.

(1) INTRODUCTION

In a previous paper,[1] the basic limitations of the single-transient high-speed oscillograph were analysed: it was shown that a time resolution of 10^{-14} sec per spot width of deflection should be practically obtainable with not unreasonable deflection sensitivity, but that, on the other hand, to make use of such high resolution, very considerable advances in the deflector-system design would be necessary. Nevertheless, by the application of the design principles evaluated, a considerable economy in design and optimization of results is possible in oscillographs designed for lower recording speeds. Previous ultra-high-speed oscillographs, such as those described in References 18–20 of the first paper, were designed with little consideration of the basic design limitation.

In order to verify some of the conclusions regarding temporal resolution, and for application in the study of certain high-speed transient phenomena, it was decided to design and construct an oscillograph to give a resolution between 10^{-12} and 10^{-13} sec with moderate deflection sensitivity. Since it was desired to have a very versatile instrument, it was decided to design it as a demountable continuously-pumped unit, and to sacrifice certain of the conditions required for optimum performance in favour of convenience and compactness. Thus, the final imaging lens operates at a magnification greater than unity instead of at a fractional magnification shown to be desirable in the analysis. In addition, the lens size is considerably less than the optimum. In spite of these limitations, the performance is very promising.

It was decided to utilize a maximum voltage of 40 kV on the electron gun, which would be of conventional design as used on the electron microscope. A final spot diameter of 20 μ was chosen as a design target, this being about the limiting resolution of the photographic plate to electrons. The electron gun using a 5-mil diameter filament wire gives a spot diameter of about 50 μ at half height of a Gaussian distribution, so that an overall demagnification of about $2\frac{1}{2}$ was required. A considerable saving in length results if a 2-lens system is used, the first giving a high demagnification and the second a small magnification.

The design is now considered with reference to the design equations given in the previous paper. An electron gun operating with theoretical best performance at 30 kV (V_0), and 3 amp/cm^2 cathode loading is assumed; the brightness [eqn. (2)] is then 110 kA/cm^2/steradian. For a resolution time of 5×10^{-13} sec/spot width, when using a photographic plate of sensitivity 4×10^{-12} coulomb/cm^2, a beam half-angle [eqn. (3)] of 0·0048 rad is required. A projection distance from final lens to plate of 20 cm was chosen as convenient for the deflector systems envisaged, and, from consideration of overall size, a final lens magnification of 3·5 was selected. The final lens focal length then becomes 4·7 cm. A suitable lens design [eqn. (6*c*)] has a gap spacing of 1·25 cm and bore diameter of 1·0 cm. The spherical aberration then has a value of about 100 cm [eqn. (6*a*)]. The required excitation is given by[2] [eqn. 6(*c*)]

$$NI_L = 5\sqrt{[V_0(s + D)/f]} \text{ ampere-turns}$$
$$= 600\,\text{AT}$$

A suitable first lens was designed, having a pole-piece spacing of 0·48 cm and bore of 0·64 cm. With 1000 AT, this gives a focal length of 0·9 cm. With the source and the lens separated by 10·5 cm, a demagnification of 11·7 is obtained, giving an overall maximum demagnification of 3·4. The overall length from the cathode to the photographic plate is 38·5 cm.

For deflector plates spaced at twice the beam diameter and 2 cm long, the deflection sensitivity would be 1 spot-width per volt assuming no transit-time effect.

(2) CONSTRUCTION

(2.1) Electron Gun

The electron gun (Fig. 1) comprises a 5-mil diameter tungsten hairpin filament as used in the type E.M.3 electron microscope.[3] The filament is spot-welded on to two nickel pins held in a

Mr. Haine is, and Mr. Jervis was formerly, in the A.E.I. Research Laboratory.
Mr. Jervis is now with the A.E.I.–John Thompson Nuclear Energy Co., Ltd.

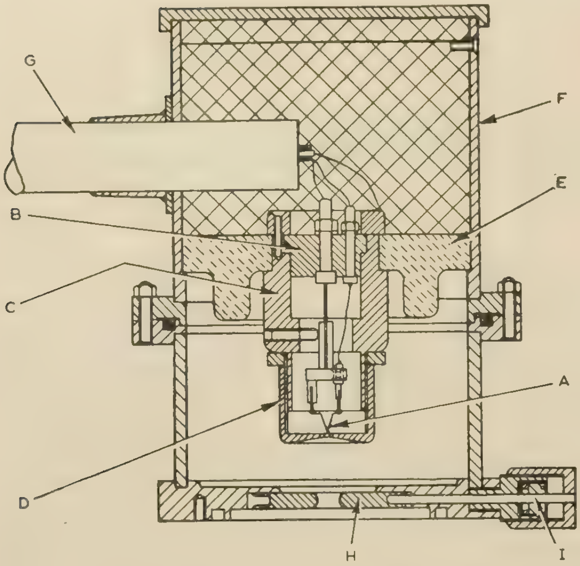


Fig. 1.—Cross-section of electron gun.

A. Filament.
B. Steatite insulator.
C. Body of cathode assembly.
D. Gun shield.
E. Porcelain insulator.
F. Outer case.
G. H.V. connection cable.
H. Anode.
I. Alignment rods.

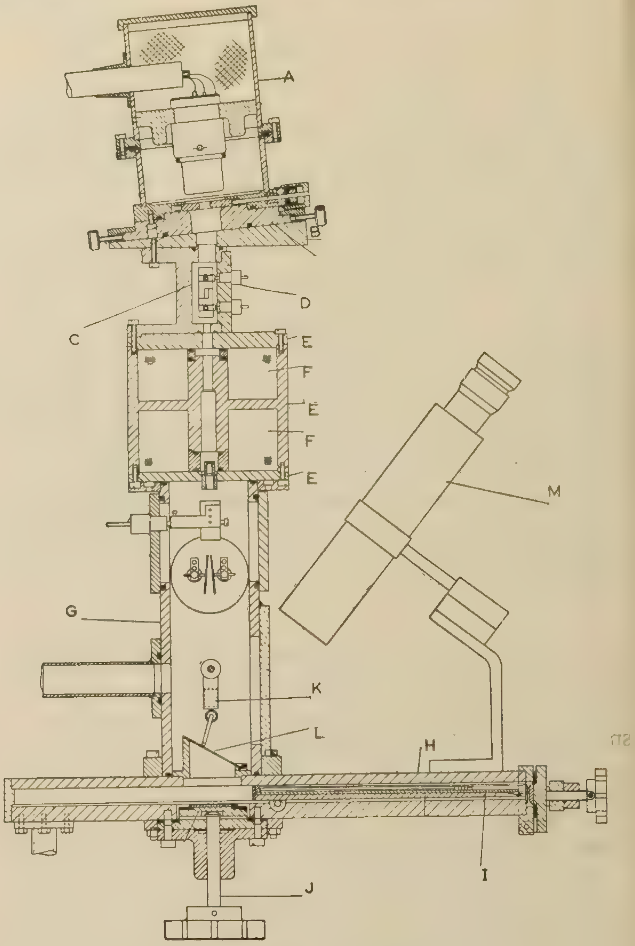


Fig. 2.—Cross-section of oscillograph.

The overall height is 24 in.

A. Electron gun.
B. Wedge plate.
C. Beam-trap chamber.
D. Insulators for beam-trap plates.
E. Lenses.
F. Coils.
G. Deflection chamber.
H. Camera box.
I. Cassette.
J. Actuating screw.
K. Faraday cage.
L. Fluorescent screen.
M. Telescope.

steatite insulator. The assembly plugs into two split-tube sockets in a copper block, mounted on the end of a spring-steel pin, one socket being insulated from the block with mica washers. The spring-steel pin is fixed at the top end in a steatite insulator. The insulated filament socket is connected to a second pin through this insulator by a flexible lead insulated with steatite beads. The filament can be centred by two perpendicular Allen screws in the body of the cathode assembly. The negatively-biased gun shield screws on to the assembly with a locking ring, allowing the relative height between filament and shield to be adjusted. The cathode assembly is insulated from the main accelerating voltage by a porcelain insulator which fits against a small step in the outer cylindrical earthed box. Connections from the high-voltage unit are made by a rubber-insulated cable with three central conductors for filament-heating current and bias. After connection, the top of the gun tube above the insulator is filled with bitumen cable-sealing compound, and a top lid is fitted. Thus, the whole external surface of the gun is earthy. The cathode unit is demountable by a rubber-sealed flange joint for replacing filaments.

The anode of the gun is a flat plate with central aperture, and is movable[4] in two horizontal directions for alignment of the emergent beam by means of two rods, oriented at right-angles, sealed by hard-rubber glands and pushed in and out by screwed thimbles.

The electron gun is mounted on a plate (Fig. 2) which slides over a rubber vacuum gasket and which can be traversed in two perpendicular directions by two screws. The wedged plate is inserted to tilt the electron gun to prevent light from the filament reaching the photographic plate. The beam is bent to the axis of the instrument by adjustment of the anode plate.

(2.2) Deflection Chamber and Lenses

The beam-trap chamber consists of a tube with end flanges cut away at one side to allow the flange to be fitted and sealed with a rubber gasket. The beam-trap plates comprise two cross-connected pairs, supported on ceramic lead-through insulators.

The two magnetic electron lenses, which are completely demountable, are made in one unit using the design principle for electron-microscope lenses described by Haine.[5] The two lenses, of axial symmetry, comprise three pieces of iron, the vacuum seals being made between the poles by double rubber gaskets. The top and bottom plates are spigoted on to the middle piece, and the three pieces of iron are machined to give the highest degree of concentricity throughout. The coils are wound (with about 10000 turns of 33 s.w.g. wire) on separate brass bobbins with pressboard insulation.

The deflection chamber comprises a square-section block of steel bored with a circular hole from end to end. Four ports are bored in the sides in two opposite pairs at different heights from the top for the insertion of the deflector plates which are mounted on metal or Perspex flanges sealed with rubber gaskets. Normally, only two of these plates are used. The other two are included to add flexibility to the instrument, e.g. to fit special

travelling-wave deflector systems which might be required with leads passing right through the instrument. The deflector systems are aligned by a jig which ensures that their axes are coincident with the beam axis.

The camera is based on that of the type E.M.3 electron microscope[3] and consists of a rectangular box blanked off at one end and fitted with a hinged gasket-sealed door at the other. The camera body is made from Duralumin, the pieces being held together and sealed by Araldite resin. A circular hole connects the camera to the deflecting chamber, to which it is sealed by a rubber gasket. A 5in × 1in photographic plate is used, this being contained in a light-tight cassette with a sliding top. After insertion in the camera, the cassette is driven forward on a carrier plate operated by a pinion and rack. The lid is held back by a simple catch, so that the plate can be exposed to take a series of oscillograms, and then, on retraction, the lid slides into position ready for removal into the light. The 5in × 1in plate conveniently accommodates six large or ten small oscillograms, their position being indicated on a scale and pointer on the top of the camera.

The camera can be isolated from the main body of the tube by a sealing plate with rubber gasket. The sealing plate is pushed against the end of the deflecting chamber tube by the actuating screw.

The tube is pumped by an air-cooled oil diffusion pump backed with a single-stage rotary pump. A flap valve is fitted on top of the diffusion pump, so that the instrument can be dismantled without cooling the diffusion pump. A by-pass line and valve to the backing pump is provided for rough pumping of camera air-lock or main tube.

All rotating joints on the tube are sealed with simple hard-rubber gland seals.

The beam current reaching the photographic plate is measured by a Faraday cage, which can be swung into the axis of the instrument on a spindle passing through a rotary vacuum seal. The current is measured either on a microammeter with the beam-trap inoperative, or with a conventional oscillograph when the beam trap is pulsed.

When the oscillograph is set up, the trace is viewed on a fluorescent screen through a lead-glass window. The viewing screen is fixed to a hinged spring-loaded flap which is raised when taking photographs, and is operated through a rotary vacuum seal. The trace, being too small to be viewed conveniently with the unaided eye, is normally viewed through a wide-aperture telescope with a magnification of 10.

The depth of focus of the oscillograph is smaller than the distance between the viewing screen and the photographic plate, and so a small correction must be made in the lens current when taking a photograph after focusing the beam on the viewing screen. The correction is made very simply by switching in a shunt across the lens just before taking a photograph.

(3) DEFLECTOR SYSTEM

The time-base deflector plates are conventional in design, being 2cm long, 1cm wide and spaced 0·3cm apart with a throw of 18cm. The deflection sensitivity is about 0·5 spot width per volt for a spot diameter of 40 μ as shown in Fig. 5(*b*).

The demountable construction of the oscillograph enables the deflector systems to be changed easily, and various Y-deflector systems have been used to suit the signal being observed. For frequencies up to about 1000Mc/s, plates similar to the X-deflector system are suitable. At higher frequencies the transit time of the electrons through the plates becomes appreciable compared with the period of the deflecting voltage, and the sensitivity decreases. At frequencies in the region of 10000Mc/s, which have been used for testing the resolution, parallel-wire deflectors were used in the manner suggested by Rudenberg,[6] the wires being 0·8mm in diameter and 3mm apart. It can be shown[7] that, with these dimensions and a 30kV electron beam, the sensitivity at 10000Mc/s is about 30% of the d.c. value. With this arrangement the maximum deflection available was found to be limited by voltage breakdown of the vacuum in the oscillograph. When greater deflection was required, systems of higher sensitivity were used. These consisted of parallel wires connected by suitable delay paths to form a travelling-wave system. The design of these systems is beyond the scope of the paper, and has been discussed elsewhere.[8,9] In Fig. 2, conventional plates are shown for both X- and Y-deflection.

(4) ELECTRICAL CIRCUITS

(4.1) Power Supplies

The high-voltage supply for the electron gun is obtained from a 12kc/s transformer feeding a copper-oxide-rectifier multiplier stack based on the design currently used in the E.M.3A and E.M.4 electron microscopes.[10] It is housed in an oil tank, 9in × 9in × 18in, and is capable of giving a maximum voltage of 50kV stabilized to about 0·1%.

The gun filament is fed from a 60kc/s oscillator through an air-core transformer mounted in a Perspex insulator terminating the high-voltage cable in the oil tank.

The lenses are fed from simple conventional stabilized supplies.

(4.2) Timing Wave and Beam-Trap System

The resolution of the oscillograph is measured from single-stroke oscillograms of sine-wave voltages, and is given by

$$\tau = \frac{1}{f_t\sqrt{(a^2\pi^2 + b^2/n^2)}}$$

If the distance between cycles is small compared with the amplitude,

$$\tau \simeq \frac{1}{a\pi f_t}$$

It can be seen that for $\tau = 10^{-12}$ and $a = 30$, a timing frequency of about 10000Mc/s is convenient.

The timing-frequency wave was obtained from a 9080Mc/s magnetron type CV2167, pulsed for 1-microsec from a radar-type modulator. The magnetron output was fed down a length of waveguide which was transformed into a concentric-line feed as shown in Fig. 3, the matching being adjusted by plungers on the concentric line and the waveguide. The concentric-line output was then converted to a push-pull output by an unbalance-balance transmission-line transformer, and connected to the Y-deflector system through a Perspex vacuum flange. No attempt was made to optimize the matching conditions, since adequate deflection was obtained with the arrangement described.

Ideally, the beam trap should release the beam for a time equal to the time-base sweep duration, which was about 2 millimicrosec. The generation of a de-trapping pulse of such a length is difficult, and since it is not necessary for temporal resolution measurements, this system was not employed, the arrangement adopted being shown in Fig. 3. The oscillograph beam was released by a 1kV 1-microsec de-trapping pulse obtained by diode clipping of the output from a resistive attenuator connected across the magnetron modulator output.

The modulator could be triggered manually or at a low frequency through a pulse generator. The oscillograph time-base was triggered through a variable-delay circuit from the

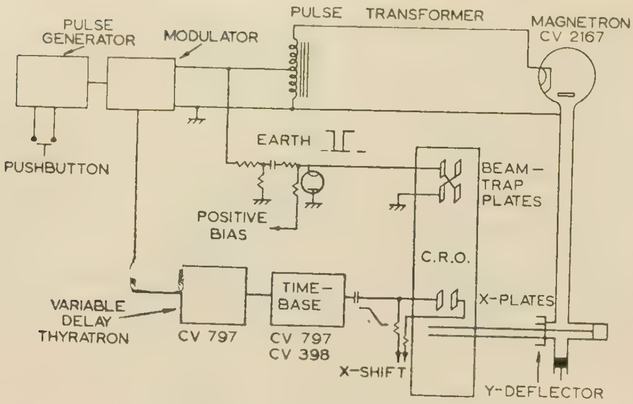


Fig. 3.—Block schematic of timing circuits.

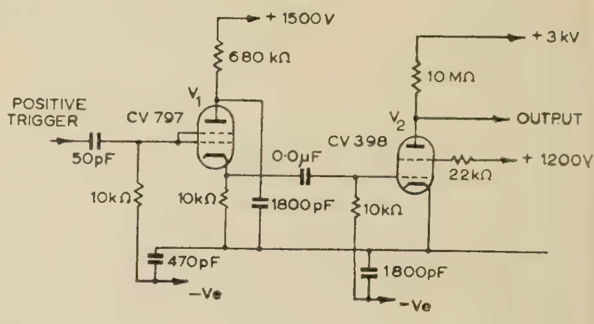


Fig. 4.—Time-base circuit.

N.B. V_1 is gas-filled.

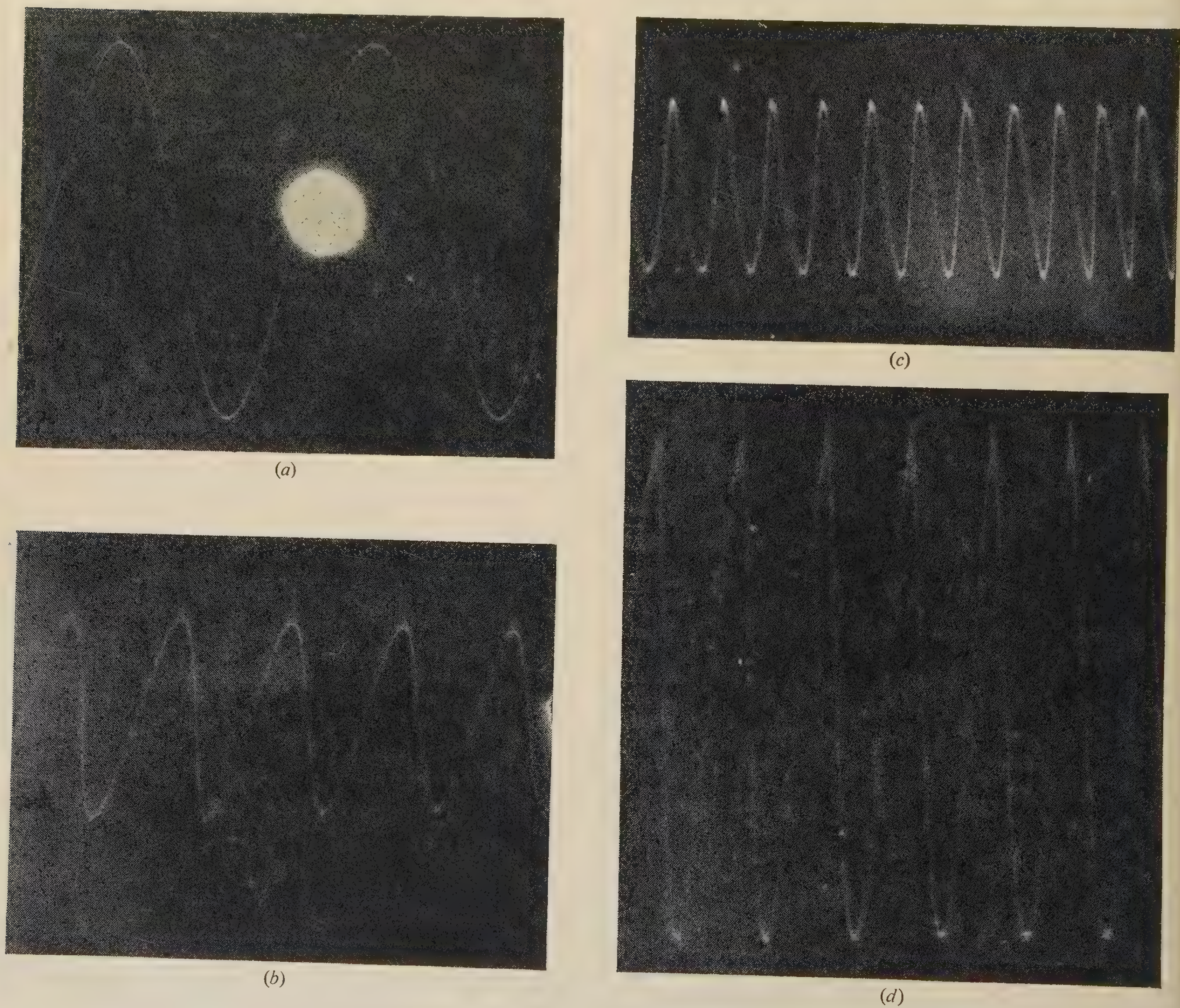


Fig. 5.—Single-transient voltage oscillograms.

(a) $f_t = 280$ Mc/s, $\tau = 2 \cdot 3 \times 10^{-12}$ sec.
(b) $f_t = 9\,080$ Mc/s, $\tau = 4 \times 10^{-13}$ sec.
(c) $f_t = 9\,080$ Mc/s, $\tau = 5 \times 10^{-13}$ sec.
(d) $f_t = 9\,080$ Mc/s, $\tau = 1 \cdot 7 \times 10^{-13}$ sec.

Owing to difficulties of reproduction, the clarity of these oscillograms is inferior to that of the original negatives.

modulator, this delay being adjusted so that the time-base operated when the magnetron pulse had reached its peak value. Since this pulse was much longer than the time-base, accurate synchronization of the time-base with the de-trapping pulse was unnecessary.

(4.3) Time Base

To make adequate use of a resolution of 5×10^{-13} sec/spot-width, a time-base with a deflection of one spot width in 5×10^{-12} sec is required. The maximum time-base speed is set by the time taken to discharge the stray capacitance of the deflector plates, leads and discharge valve when passing the peak current of the valve.

The plates described in Section 3 have a sensitivity of 0·5 spot-width/volt, so that for a speed of 5×10^{-12} sec/spot-width the required rate of change of voltage is 4×10^{11} volts/sec.

A number of time-base arrangements were considered, the one finally adopted being a modification of that originated by Fletcher,[11] and more fully described elsewhere.[12] The grid of the tetrode discharge valve V_2 (Fig. 4) is held cut off by a large negative voltage. The anode and screen grid are fed through high resistances from a high-voltage supply. A positive driving pulse is applied to the grid, and the anode-to-earth capacitance is discharged by the anode current. Since the tetrode is operated above the kink in the I_A/V_A characteristic, a fairly constant discharge current is obtained, so giving a reasonably linear sweep.

The time-base speed is given by $dV/dt = I_d/C$, so that, for $C = 20$ pF and $dV/dt = 5 \times 10^{11}$ volts/sec, $I_d = 10$ amp; this current is obtained from the discharge valve V_2, which is a type 715 (CV398) pulse tetrode. The driving pulse applied to its grid is obtained from a 20A3 (CV797) tetrode thyratron, V_1. This valve is normally cut off and, when triggered with a positive pulse, discharges the anode capacitance into the cathode circuit, so driving the grid of the discharge valve hard positive.

The speed of the time-base is illustrated by the oscillogram shown in Fig. 5(*c*). The frequency of the signal displayed is 9080 Mc/s, and the voltage corresponding to the distance between peaks of the wave is 60 volts, so that the speed is $5\cdot5 \times 10^{11}$ volts/sec. The maximum sweep voltage available is about 1·5 kV.

(5) RESULTS

The resolution of the oscillograph is obtained from oscillograms of the type illustrated in Fig. 5. Since the photographic process can alter the apparent spot size and density, actual measurements are taken direct from the photographic plate with a travelling microscope.

The quality of lower-frequency oscillograms (280 Mc/s) is shown in Fig. 5(*a*), which was taken without the beam tilt in operation and thus shows the light from the filament. The spot width is about 30 μ, the peak-to-peak amplitude being about 500 spot-widths. The oscillogram shows some astigmatism and deflection defocusing, and attempts to reduce this are being made.

Figs. 5(*b*)–5(*d*) show oscillograms taken at 9080 Mc/s. The following data relate to Fig. 5(*b*), which is typical of a large number:

Frequency of timing wave	9080 Mc/s
Peak-to-peak trace amplitude	90 spot widths
Temporal resolution	4×10^{-13} sec
Electron accelerating voltage	40 kV
Spot diameter at centre of trace	40 μ
Collected current	135 μA
Current density in spot	10·8 amp/cm^2
Aperture diameter	0·27 cm
Aperture semi-angle	0·0055 rad
Gun brightness	10^5 amp/cm^2/steradian

The temporal resolution is calculated from[1]

$$\tau = \frac{\sigma_d}{J} = \frac{1\cdot2 \times 10^{-7}}{V_0 J} \simeq 3 \times 10^{-13}\,\text{sec}$$

for $V_0 = 40$ kV and $J = 10$ amp/cm^2. This is in fair agreement with the experimental value. Fig. 5(*c*) shows a similar oscillogram giving inferior resolution but greater density, and for Fig. 5(*d*) the deflection was increased, giving a resolution of $1\cdot7 \times 10^{-13}$ sec. Some difficulty was experienced in reducing coupling between the X- and Y-deflection systems at microwave frequencies: this causes a distortion of the time scale but does not seriously affect resolution measurements.

Oscillograms taken with amplitudes larger than that in Fig. 5(*d*) had a resolution of 10^{-13} sec with a correspondingly lower density; although visible on the negative, they are unsuitable for enlargement. The resolution of the oscillograph is thus indicated to be about 2×10^{-13} sec.

(6) CONCLUSIONS

The oscillograph described has a resolution of about 2×10^{-13} sec, which is in agreement with that to be expected from the theoretical design.

To exploit fully such writing speeds in the recording of transients, the development of suitable travelling-wave deflector systems would be necessary, but there appears to be no fundamental reason why this should not be successfully achieved. It is probable that many, if not most, applications might require special deflectors to suit particular conditions. The use of a demountable tube has important advantages in this respect. With modern demountable vacuum technique, such an instrument can be made fully reliable and simply maintained.

It should be emphasized that the oscillograms shown were taken mainly for the purpose of measuring the temporal resolution of the oscillograph. The defocusing, particularly at the extremes of deflection, is due partly to astigmatism but largely to cross-coupling effects. There seems no fundamental reason why these should not be reduced, with a resultant improvement in the accuracy and quality of the oscillograms.

(7) ACKNOWLEDGMENTS

The authors are pleased to acknowledge the assistance given by Mr. R. T. Taylor in the experimental work connected with the oscillograph. Thanks are also due to Dr. T. E. Allibone, Director of the Laboratory, for permission to publish the paper.

(8) REFERENCES

(1) HAINE, M. E., and JERVIS, M. W.: 'The Ultimate Performance of the Single-Trace High-Speed Oscillograph', *Proceedings I.E.E.*, Paper No. 2306 M, February, 1957.

(2) LIEBMANN, G.: 'A Unified Representation of Magnetic Electron Lens Properties', *Proceedings of the Physical Society*, 1955, **68B**, p. 737.

(3) HAINE, M. E., PAGE, R. S., and GARFITT, R. G.: 'A Three-Stage Electron Microscope with Stereographic Dark Field and Electron Diffraction Capabilities', *Journal of Applied Physics*, 1950, **21**, p. 173.

(4) HAINE, M. E.: 'Improvements Relating to the Construction of Electron Guns', British Patent Application No. 146541, April, 1952.

(5) HAINE, M. E.: 'Some Simplified Magnetic Lens Design Features', International Conference of Electron Microscopy, London, 1954.

(6) RUDENBERG, H. G.: 'Deflection Sensitivity of Parallel-Wire Lines in Cathode-Ray Oscillographs', *Journal of Applied Physics*, 1945, **16**, p. 279.

(7) JACKSON, B.: 'The Recording of High-Speed Electric Transient Phenomena', M.Sc. Thesis, Manchester University, 1953.

(8) LEWIS, I. A. D., and WELLS, F. H.: 'Millimicrosecond Pulse Techniques' (Pergamon Press, London, 1954).

(9) TALBOT, R. V., and JOHNSON, L. M.: 'The TW-10: A High Writing Speed Cathode Ray Tube with Distributed Deflection' (Naval Research Laboratory, Washington, U.S.A., May, 1954), NRL Report 4377.

(10) PAGE, R. S.: 'A Compact Console-Type Electron Microscope', *Journal of Scientific Instruments*, 1954, **31**, p. 200.

(11) FLETCHER, R. C.: 'Production and Measurement of Ultra High-Speed Impulses', *Review of Scientific Instruments*, 1949, **20**, p. 861.

(12) JERVIS, M. W., and TAYLOR, R. T.: 'A Millimicrosecond Time Base', *Electronic Engineering*, 1957, **29**, p. 218.

DISCUSSION BEFORE THE JOINT MEETING OF THE MEASUREMENT AND CONTROL SECTION AND THE RADIO AND TELECOMMUNICATION SECTION, 12TH FEBRUARY, 1957

Mr. D. F. Oakeshott: To a large extent the speed of these instruments is not at the present matched by the significance of the results, at any rate when dealing with single transients. By this I mean that it is one thing to sweep a beam across a plate or screen at these very high speeds—albeit a clearly visible beam—and quite another to persuade it to move in a manner which faithfully reflects the initial voltage or current phenomenon. The authors do not disregard this, of course, and, in fact, refer briefly to the travelling-wave deflector system, etc., and the formidable difficulties of its application under many circumstances; but it nevertheless remains the salient feature of this work. It is probably not without significance that the oscillograms reproduced are of a semi-continuous sine wave. Voltage division and the design of current shunts for single transients at these frequencies are very difficult matters.

The second point is not, perhaps, so fundamental, and concerns the comparison of sealed-off and continuously-pumped tubes. Although the authors state that the two methods are comparable, they have, in fact, used a continuously-pumped tube themselves, largely because it makes the construction and development a relatively straightforward matter. But it would be useful to know which they prefer in the end-product, first for the most advanced oscillography and, secondly, for normal laboratory impulse techniques. One feature of special importance here is the ratio between the Y-plate deflection and main acceleration voltages. In the lower-voltage tubes one often encounters the condition where the Y-plate voltage not only deflects the beam but appreciably accelerates or decelerates it, thus altering the sensitivity. This imposes certain limitations on the method of application of calibration and phenomenon voltages. This effect is virtually absent, one presumes, in the authors' own oscillograph, with its 50 kV beam, and I should like the authors' comment on this point.

In Section 8 of Paper No. 2306 the authors state that 'for the highest speeds and highest deflection sensitivities shown in Table 2, the tube length is becoming rather excessive'. It seems to me that only one of these two parameters is increasing with length, according to Table 2.

The consideration of the limits of indirect recording is necessarily a little brief, and I wonder whether the authors are satisfied that all that can be done with current techniques has been done regarding the choice of film and developers. There is evidence that the pyro-soda developer, for example, gives effectively higher resolution than some so-called fine-grain ones.

Very brief mention is made of the deflection defocusing change in the oscillograph, which the authors hope to eliminate. I should be interested to know what steps they have in mind, for it is a difficult matter.

Finally, will the authors elucidate the bending of the beam to the axis of the instrument with the anode plate? From the illustration it seems that one can slide it only in a plane normally to the gun axis, and it would seem that this will merely act as a blocking diaphragm, so that to get the beam axial the light would also be axial.

Dr. R. Feinberg: In the analysis of the electron beam system, only temporal resolution and deflection sensitivity have been discussed: these two aspects are primary criteria. On the other hand, deflection defocusing and the transit-time effects are secondary criteria, but they are important when considering the ultimate design. For example, the width of the electron beam in the deflection field affects the degree of deflection defocusing, a small beam diameter giving a small deflection defocusing, and vice versa. Thus the design of the electron lens system must be considered, not only from the aspect of spot size, but also with regard to the beam width in the deflection field.

The term 'brightness' is not a happy choice with reference to the electron gun system: a radiometric term would be more appropriate than a photometric one.

In comparing the direct photographic method, used in demountable continuously-pumped oscillographs, with the indirect photographic method, used with sealed-off cathode-ray tubes, it is important to be aware of the present limit in our knowledge of the performances of phosphors for high-speed sealed-off oscillograph cathode-ray tubes and of high-speed photography. The excitation time of the phosphor is a fraction of a micromicrosecond, and at the moment there is no information available on the relevant phosphor performance or on any failure of the reciprocity law in photography. Hence it is too early to make a definite statement on the ultimate performance of a sealed-off oscillograph tube, particularly also in view of new and promising techniques of screen laying.

The figure given for the phosphor saturation—a few milliamperes per square centimetre at the most—appears to refer to a low-electron-velocity excitation. The higher the electron velocity the deeper the electrons penetrate into the phosphor layer; therefore, the higher will be the current density at saturation, which is a volume effect but not an area effect. Measurements, for example, on flying-spot scanner tubes with an accelerator voltage of 25 kV indicate a saturation at a figure above 1 amp/cm^2.

Paper No. 2307 does not indicate the signal voltage at which the insulation broke down in vacuum. We tested a sealed-off tube with wire deflectors of 1 mm diameter spaced 4 mm at the centres, with a 9 800 Mc/s voltage of well above 6 kV peak before breakdown, which occurred externally along the glass surface between the electrode seals.

Also, I should like to know the deflection sensitivity which was obtained but which is not stated.

Dr. R. F. Saxe: Up to now the oscillographs which have been used for recording have been unable to respond in a time less than about 10^{-9} sec. The authors have extended this limit down to 10^{-12} sec, and now the problem is one of recording such rapid phenomena. One of the difficulties is that of putting the transient into the oscillograph, i.e. coupling it with the

electron beam, without distortion; it is one thing to put in a sine wave and it is another to put in a transient of unknown shape, look at the oscillogram and be sure that the trace is a true reproduction of the transient. About the worst one can do with a sine wave is to reduce its amplitude.

I have some transients of the order of 10^{-10} sec, but no oscillograph to display them. However, I have seen elsewhere a sealed-off tube, or an attempt at a sealed-off tube, with travelling-wave deflector systems, which has endeavoured to get down to 10^{-10} sec. It is not promising enough to make me go out and buy one.

This work reintroduces the problem of the time-base. If we accept a spot width of the order of 20 microns, and a scan length of 1 cm, we need a time-base of 5×10^{-10} sec. This can possibly be done reasonably with some brute force, but if our transient is beyond control in its time of initiation, we are in a difficulty, and unless we are not to grow tired of waiting we must ensure that initiation jitter is less than 5×10^{-10} sec. This is not easy. We must ensure that our time-base starts within a reasonable time of the trigger transient, which might have an awkward size or shape. If it takes a long time to commence we must delay our signal, and this again may introduce distortion, which it is not easy to overcome or even to estimate if we have a peculiarly shaped transient.

Mr. H. G. Riddlestone: I should like to emphasize that the difficulty which we experience in using a high-speed oscillograph is not lack of resolving power but difficulty in obtaining faithful reproduction on the screen of the transients applied to the plates. We have seen that it is possible theoretically and practically to record 10 000 Mc/s. Immediately we consider the deflecting system we think in terms of thousands of megacycles per second, and with amplifiers we come down another order. If attention could now be given to this part of the oscillograph with the same success, great progress could be made.

However, in addition to showing that resolution can be considerably improved, the authors have pointed out that the resolution of our present tubes could still be obtained with smaller tubes. If this can be done, the problem of connection to the deflecting system might be eased, so that the authors' analysis has not taken us out of the range of practical use.

On the question of sealed-off or continuously-evacuated tubes, the possibility of changing the deflection system of the latter for a particular application is an important point to consider. On the other hand, the convenience of the sealed-off tube cannot be lightly disregarded. Could the sensitivity of the indirect method of photography be improved to take greater advantage of the greater inherent sensitivity? Also, have the authors any quantitative information on the difference between the aberration of the negative-grid gun compared with the beam-trap system, and have they any observations to make on why they chose the magnetic lens rather than an electrostatic system?

Mr. P. A. Einstein: Phosphor saturation is certainly a volume effect, in so far as it can be considered in terms of the number of unenergized luminescent centres available at any given time and the rate of arrival of electrons. Thus, at high steady current densities, many of the electrons are unable to find unenergized luminescent centres and consequently the phosphor efficiency (luminous power/electrical power) is reduced. With short pulses, the onset of this saturation might be expected to occur for higher current densities than in the case of steady irradiation, particularly when the pulse length is of the order of, or less than, the natural rise time of the phosphor.

Dr. Feinberg points to the effect of the increasing depth of penetration with electron voltage; this would be expected to lead to similar delay in the onset of saturation inasmuch as the electrons now encounter a greater effective volume of phosphor. There is evidence that saturation decreases with increasing voltage.*

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. M. E. Haine and **M. W. Jervis** (*in reply*): Several speakers refer to the difficulty of coupling the input transient to the electron beam in a way which gives a faithful reproduction of the input waveform. We agree that, now an oscillograph with a high writing speed is available, the main problem remaining is that of the deflecting system. This requires an investigation of a magnitude similar to that of obtaining a high writing speed. The problems of voltage division and current shunts are considerable, but are receiving attention by several workers, as reported in References 1 and 2 of Paper No. 2306 M, but at present the difficulties of amplification at high frequencies limit oscillographs to the display of relatively large voltages. It must be emphasized that, although the analysis has been applied to deduce the limiting performance which could be obtained with conventional electron guns and lenses, the results are just as readily applied to design the optimum tube for any lesser performance. Such application shows that considerable economy could be achieved even at moderate performance. The most important application is possibly to give the highest deflection sensitivity for a given time resolution.

In reply to Mr. Oakeshott, the sealed-off instrument is undoubtedly preferable for most applications. The demountable tube has special advantages in that special or interchangeable deflectors can be used; in general, a finer trace is obtainable. The simplicity of modern demountable vacuum technique makes it possible for the demountable instrument to be quite light and portable.

There is considerable controversy between the manufacturers and the users of photographic materials regarding the advantages of various processing schemes. Special techniques were not used in the work described in the papers, since it was desirable to arrange conditions resembling those existing in the investigation of Digby *et al.* (Reference 9). With reference to calibration errors, the change of sensitivity caused by transit time is usually greater than that caused by other effects. Transit-time errors can be calculated, but only rather approximately, because of unknown edge effects.

It is true that deflection defocusing can ultimately limit an oscillograph performance. However, the use of optimum design leads to minimum beam angles, which minimizes defocusing. The deflection defocusing obtained on our instrument resulted partly from pick-up of the signal and partly because excessively large deflections were used to obtain sufficient deflection speeds to test the limit of recording sensitivity. At low speeds and deflections of 100–200 spot widths, no deflection defocusing was observed on low-speed traces.

Mr. Oakeshott remarks that the instrument throw depends on deflection sensitivity but not the time resolution. This is not strictly so, since the time resolution depends on the deflection sensitivity also.

The deflection of the beam by the anode plate results from the fact that the electrostatic field in the aperture produces a divergent lens. The act of moving this lens off centre produces the deflection.

In reply to Dr. Saxe, distortion can occur even with sine waves, owing to coupling between the time-base and the signal

* Dowling, P. H., and Sewell, J. R.: *Journal of the Electrochemical Society*, 1953, 100, p. 22.

deflectors. One way of avoiding the difficulty of synchronizing the time-base with the signal is to take part of it and distort it into a suitable sweep voltage. The actual signal is delayed for a suitable period. Such a system was demonstrated by the Radar Research Establishment at the Physical Society's Exhibition in 1956.

With reference to Mr. Riddlestone's remarks, there is no doubt that optimum design should appreciably reduce tube sizes and help to reduce the difficulties of producing suitable deflection systems. There seems no reason why a suitable electron gun using an oxide cathode should not be designed. The conventional type of gun is capable of appreciable improvement. The use of a negative grid instead of a beam trap offers some advantage in simplicity.

We have used magnetic lenses in our experimental instrument partly because we have more extensive data on them, and partly because we consider them more convenient in a demountable instrument. The spherical aberration for an electrostatic lens of comparable dimensions is little different from that for a magnetic one. For a sealed-off instrument there are still advantages in the use of a magnetic lens, in that it can be arranged outside the vacuum envelope, allowing the use of a bigger lens with corresponding reduction of spherical aberration.

We thank Dr. Feinberg and Mr. Einstein for their comments on phosphors. Dr. Feinberg is incorrect in assuming that the phosphor is less efficient for excitation times less than the build-up time; on the contrary, there is every reason to suppose that the phosphor saturation is less for such short excitation times. The efficiency of phosphors is already so high that little improvement can be expected from improved efficiency. In any case as has been pointed out in the papers, the recording is already very near the limit set by quantum noise in the electron beam.

The deflection sensitivity of the parallel-wire system is 0·05 spot width per volt and the breakdown occurred at 10 000 Mc/s for a voltage in the region of 5 kV, the actual value depending on the gas pressure in the oscillograph.

DISCUSSION ON 'THE CRYSTAL PALACE TELEVISION TRANSMITTING STATION'*

Before the South Midland Centre *at* Birmingham *3rd December, and the* Northern Ireland Centre *at* Belfast *11th December*, 1956.

Mr. W. H. Brent (*at Birmingham*): The opening of the new station will not improve reception in the area around Huntingdon, since the gain in power is offset by the move to the south, and therefore the Post Office can still do no more than sympathize and give advice on interference problems, the field strength being still very low. In the Northampton area, however, where the contours of the Sutton Coldfield station overlap those of the London station, there will evidently be some improvement from the latter when full power is radiated. Some cases of interference, such as pattern interference from the local oscillator of a Channel I receiver on a neighbouring Channel IV receiver, might be overcome by persuading the owner of the latter that he would get better reception by changing over to Crystal Palace.

On the matter of security of transmission, the dual transmitter arrangement is certainly an ingenious method, when it is recalled that Sutton Coldfield opened without standby, and it is one which might be considered in other transmission paths as an alternative to switching from one repeater to another. An example is to be found in the transatlantic-telephone repeaters. It was noted, however, that the Diesel-driven alternator did not maintain the station if the mains supply failed completely, and might not one of the dual transmitters have been included?

The efficient layout of the building shows the hand of the engineer rather than the architect, but the sinking of the building in the ground cannot have been lightly undertaken, on account of the cost. The staff must work under artificial lighting and with forced ventilation. Where is the air-conditioning plant housed? This might be expected to take up almost as much space as the operational plant, and presumably must be on a lower floor.

Mr. H. B. Mellor (*at Birmingham*): I notice that the parabolic reflector aerials are positioned by remote control through selsyns. Is this system also used to trim the transmitting dipoles, or is this unnecessary after the initial setting up during installation?

Mr. L. H. Tolley (*at Birmingham*): Duplicate power supplies by separate cables fed from independent networks have been used for large telephone exchanges and safeguard the supply very well. Power cables, like other plant, do fail occasionally, but I cannot recall a case when both cables have been out of service together. With duplicate power supplies, and, in effect, duplicate paths through the transmitters, and duplicate feeding to the aerial system, it is clear that the maintenance of service from the Crystal Palace station has been very carefully considered. Are the switches, which seem to be the only common points, also in duplicate?

Although there is a reference to ice on the tower, I assume that de-icing equipment is provided for the aerials. Surely the characteristics would be altered if there were a build-up of ice or snow.

I do not understand the reason for the statement that under breakdown conditions the reduction of power would be 10 dB at regional stations, unless if the normal transmitter fails at a regional station a lower-power transmitter and standby aerial are used. With reference to the reserve aerial at Sutton Coldfield, is there an appreciable shadow from the main mast and any appreciable reflection from it when the standby aerial is in use?

Mr. J. M. W. McBride (*at Belfast*): I am impressed by the

* McLean, F. C., Thomas, A. N., and Rowden, R. A.: Paper No. 2069 R, March, 1956 see 103 B, p. 633).

considerable extra capital cost which must have been involved in placing the transmitter building below ground level, and in erecting a self-supporting tower in place of the more usual stayed mast, through lack of space. Was there no possibility of obtaining a larger site, with room to erect a stayed mast and to erect the transmitter building above ground? The Shooters Hill site gives almost identical coverage, according to the theoretical results in Table 1, but no details are given of the nature of this site and no experimental field-strength contours are plotted for a transmitter at this location. Would this site have satisfied the requirements I have mentioned?

The use of duplicate feeders and distributors to the two halves of the aerial array is a good feature, but it would seem preferable from the aspect of aerial gain under fault conditions to arrange each feeder to supply two tiers of the upper half and two of the lower half, rather than as described in the paper. This alteration would more nearly preserve the aerial aperture and mean height, and would presumably give an increased gain.

Was the decision to use skeleton, or cage, dipole elements in preference to solid surfaces of revolution made because of the low wind loading of the former? The latter alternative would seem to offer better broad-band impedance characteristics and so might have allowed simpler feeding arrangements.

I should also be interested to know whether any trouble has been experienced with cross-modulation between sound and vision due to excitation of the tower structure. The resulting radiation of sum and difference frequencies might cause interference to services using the appropriate channels. Have any measurements of this kind been made?

Messrs. F. C. McLean, A. N. Thomas and R. A. Rowden (*in reply*): For convenience we have grouped our replies under speakers' names.

Mr. Brent.—Since duplicate electrical supplies from different sources are fed to the station, the risk of its being out of action for any length of time owing to mains failure was felt to be very small. The risk of failure of either supply was also considered to be so small that the expense of automatic change-over facilities in the event of failure of one of the supplies was considered to be unjustified. However, the low cost of providing a Diesel-driven alternator to provide emergency lighting facilities and to permit operation of the receiving equipment was acceptable. Because of the reliability of the mains supply, the expense of providing sufficient capacity from the Diesel-alternator set to run one of the dual transmitters was also considered to be unwarranted.

The air-conditioning plant is housed in a second storey in the front of the building, above the office block. The plant is installed at either end of this storey (input and output) and occupies about 400 ft^2 of the total area, which is much less than the space occupied by the transmitter operational plant. This leaves a considerable area of the second storey available for cooling plant in the event of further expansion.

Mr. Mellor.—The horizontal radiation pattern of the Band I transmitting aerial is not so critical that any orientation of the aerials is needed after the initial setting-up during installation. For this reason it is not necessary to trim the dipoles by remote control, as is done for the parabolic reflector aerials.

Mr. Tolley.—The r.f. switches are not in duplicate, but the failure of any switch is equivalent to the failure of one main component, e.g. combining unit, main feeder, etc. Thus the system is not at the mercy of a single switch.

De-icing equipment is being provided for the upper four tiers of the aerial. Experience makes it doubtful whether this equipment is really necessary, and it is not being fitted to the lower half of the aerial initially. Should it prove necessary to add de-icing to the lower half, the design permits this to be accomplished quite easily.

If the main transmitter at a regional high-power station fails, the standby transmitter is so arranged that it can be connected to the main aerial. The standby transmitter is 5 kW peak white as against the 50 kW peak white of the main transmitter, hence the drop of 10 dB.

The reserve aerial on the 150 ft subsidiary mast at Sutton Coldfield is not in use, having been superseded by an omnidirectional reserve aerial at the 470 ft level on the main mast.

Mr. McBride.—The site at Shooters Hill which might have become available to us was not of sufficient area to allow the erection of a stayed mast. The additional cost of building at the Crystal Palace was accepted because the service contours from this station fitted best with those of the other existing and proposed stations.

Provided that the tiers of an aerial are spaced one wavelength or more apart, the power gain (ratio) is very nearly equal to the number of tiers. Thus four tiers spaced over an aperture of eight wavelengths will give no more gain than four tiers spaced over an aperture of four wavelengths. Thus the arrangement suggested would give no more gain and would result in added complication in the distribution feeder arrangement.

Skeleton dipole elements were used to reduce the wind load; solid surfaces of revolution would not have given any improvement in broad-band impedance characteristic. It is true to say that a skeleton structure has substantially the same bandwidth as the surface of revolution on which the wires forming the skeleton fall.

No trouble has been experienced with cross-modulation between sound and vision due to excitation of the tower structure. There is no doubt that sum-and-difference frequencies are generated, owing to non-linear contacts in the tower structure, but, so far, no serious troubles have arisen due to this cause.

621.314.7

Paper No. 2367 E
July 1957

A STUDY OF HIGH-SPEED AVALANCHE TRANSISTORS

By J. R. A. BEALE, B.A., W. L. STEPHENSON, B.Sc., and E. WOLFENDALE, B.Sc., Associate Member.

(*The paper was first received* 12*th October*, 1956, *and in revised form* 4*th February*, 1957.)

SUMMARY

The discovery of an extremely fast relaxation oscillation in a junction transistor led to a study of this mode of operation. It was found that the high speed of operation was due to a combination of multiplication and punch-through.

The paper describes the static and dynamic properties of transistors which operate in this mode, and discusses the design and application of such transistors.

(1) INTRODUCTION

During an investigation into the characteristics of germanium *p–n–p* junction transistors at voltages very much greater than the normal operating voltage (the avalanche region), it was discovered that certain transistors behaved like 2-terminal negative resistances. These transistors, when connected in the circuit of Fig. 1 with the base and emitter connected together to earth,

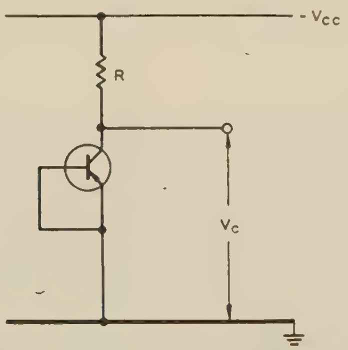


Fig. 1.—The junction transistor as a two-terminal negative resistance operating as a relaxation oscillator.

exhibited a relaxation oscillation above a certain critical value of the supply voltage V_{cc}. The waveforms of collector voltage and emitter current (shown in Fig. 2) were similar to those of a gas-discharge tube operating in a similar mode. The capacitance at the collector charges towards the supply potential V_{cc} through the load resistor R. The transistor is held in the cut-off state by the short-circuit between emitter and base until at a certain value of collector voltage the transistor conducts, and the capacitance is rapidly discharged through the transistor. The transistor then cuts off, and the cycle is repeated.

One of the transistors used was an experimental high-frequency junction transistor with an α cut-off frequency of 8 Mc/s. The waveforms obtained with this transistor were examined using a high-speed oscillograph. The rise-time at the collector was found to be less than 0·01 microsec. The amplitude of the collector voltage waveform was approximately equal to the

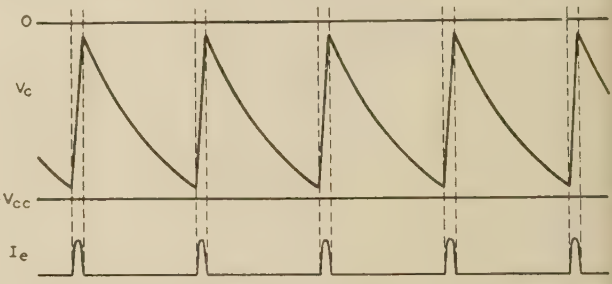


Fig. 2.—The collector voltage and emitter current waveforms of the circuit of Fig. 1.

supply voltage of 20 volts, and the amplitude of the emitter current pulse was about 50 mA. The period of oscillation was 0·1 microsec.

As the speed of operation was of an order of magnitude higher than could be achieved by using the same transistor in conventional circuits, this mode of operation was investigated.*

The investigation produced results which have led to a qualitative explanation of this mode of operation, and the paper outlines the theory and briefly describes experiments which support it. The application of these transistors is then discussed, and a description is given of circuits using transistors which were designed to operate in this mode.

(2) THE JUNCTION-TRANSISTOR STATIC CHARACTERISTICS AT HIGH VOLTAGES

If the collector characteristics of a junction transistor are examined, three turnover voltages may be found, as shown in Fig. 3: one with the emitter open-circuited (V_B or V_z), one with the base open-circuited (V_x), and another with the base connected to the emitter. This last turnover voltage may be the same as

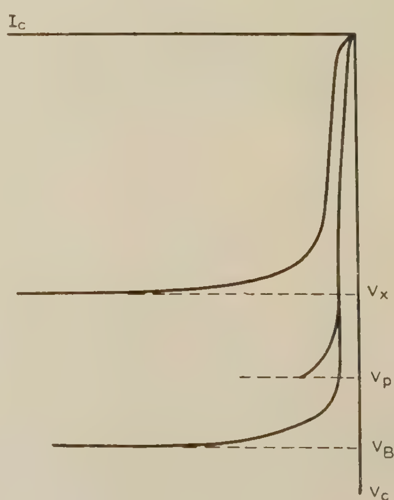


Fig. 3.—The three possible turnover voltages in a junction transistor.

* Work had been carried out in the United States on other aspects of operation in the avalanche region. The published results of this work[5,6,9,10] became available in this country during the course of the investigation and were a considerable help to the authors.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The authors are at the Mullard Research Laboratories.

that with open-circuited emitter, which will be discussed in Section 4.2.

The turnover voltage with open-circuited emitter is caused by either Zener breakdown or avalanche multiplication.[1, 2] The results of Knott, Colson and Young[3] indicate that when the resistivity of the n-type base material is greater than 0·5 ohm-cm, the turnover voltage is caused by avalanche multiplication, but below this value of resistivity the turnover is caused by Zener breakdown.

The avalanche multiplication is caused by ionization by collision in the collector depletion layer. Ionization produces hole-electron pairs, the electrons returning to the base and the holes passing into the collector. Every hole that enters the depletion layer produces on the average M–1 further holes, where M is the multiplication factor. At low current densities M is given by the expression

$$M = \frac{1}{1-(V_c/V_B)^n} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where V_c = Collector voltage.
V_B = Avalanche turnover voltage.
n = A power which is dependent on the material used ($\simeq 3$ for a germanium p–n–p transistor).

In the common-base circuit the hole current from the emitter is multiplied by the factor M in the depletion layer, and the effective current gain, α_{eff}, is given by

$$\alpha_{eff} = M\alpha_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where α_0 is the ratio of the current entering the collector depletion layer to the emitter current. As the collector voltage is raised, the factor M increases, thus increasing the effective current gain.

The lower turnover voltage with open-circuited base occurs when the effective current gain in earthed emitter, α_0', becomes infinite:

$$\alpha_{eff}' = \frac{M\alpha_0}{1-M\alpha_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

When $M\alpha_0$ approaches unity, α_{eff}' approaches infinity.

The effective current gains for a transistor with a base resistivity greater than 0·5 ohm-cm are shown as a function of collector voltage in Fig. 4.[4, 5]

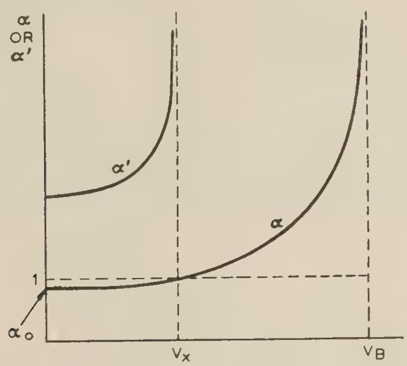


Fig. 4.—The variation of α and α' with collector voltage.

(3) THE AVALANCHE REGION AND THE RELAXATION OSCILLATION

If the transistor is operated at collector voltages greater than the turnover voltage with open-circuited base, it is operating in a region where the current gain is greater than unity, and therefore behaves like a point-contact transistor. A number of circuits have been published showing how the transistor may be used in this region.[6]

These circuits utilize an external base resistance to provide regeneration, and it was found that a number of transistors which would not operate in the circuit of Fig. 1 would exhibit a relaxation oscillation, provided that an external base resistor and an external collector capacitor were included in the circuit, as shown in Fig. 5.

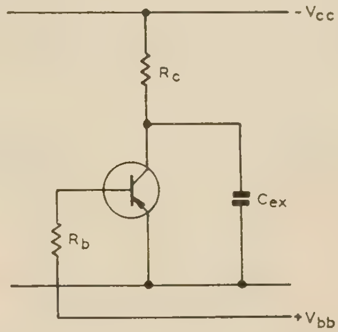


Fig. 5.—The avalanche transistor operating as a relaxation oscillator.

The start of the relaxation oscillation can now be explained. As the collector voltage is increased, so the effective current gain is increased by the multiplication in the collector depletion layer. When the collector voltage approaches V_B the collector cut-off current flowing between collector and base increases rapidly. This current flowing through the base resistor eventually increases the internal emitter-to-base potential so that the emitter starts to conduct.

The transistor now behaves like a point-contact transistor with a current gain very much greater than unity, and under normal conditions the collector voltage would fall approximately to the lower turnover voltage V_x, where the current gain is reduced to unity.

Three points about the relaxation oscillation are not explained by the above theory. First, the speed of operation is much greater than would be expected from the measured cut-off frequency of the transistor. Secondly, some transistors exhibit the relaxation oscillation without the use of an external base resistor, even though the normal internal base resistance is too small to cause regeneration. Thirdly, the collector voltage continues to fall rapidly from V_x to almost zero voltage. These points will now be considered.

(4) PHYSICAL EXPLANATION OF THE RELAXATION OSCILLATION

(4.1) The High Speed of Operation

The rapidity of the build-up of current suggests that the transit time must be shorter than when the transistor is operating normally. The normal transit time is the average time for a hole to diffuse from emitter to collector, and is given by

$$t = \frac{1\cdot 2}{2\pi f_{c0}}$$

and

$$f_{c0} \propto \frac{1}{w_b^2}$$

f_{c0} being the α cut-off frequency and w_b the base width.[7]

If the holes are aided by a field, the transit time is much less for the same base width, so that the rapid build-up of current could be explained if there were a field in the base. The most likely source of a field seemed to be the high collector voltage,

since the collector depletion layer stretches into the base in a way that has been described by Schenkel and Statz.[8] For the sake of completeness a similar description is given below.

When a p–n junction is biased in reverse, the regions on either side of the junction become depleted of mobile carriers, and in the simplest case there is a complete absence of all electrons and holes from both regions to a depth sufficient to neutralize the applied voltage. If the regions are homogeneous, as in simple alloyed junctions, the charge densities on either side of the junction are effectively constant to a certain depth, and therefore the field distribution is triangular (Fig. 6). The slopes AB,

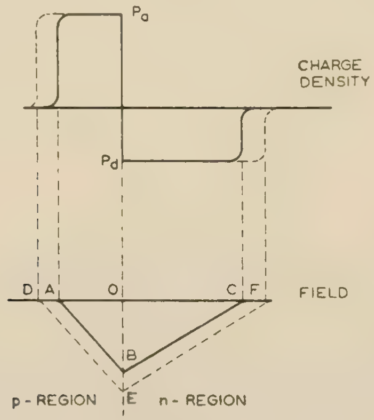


Fig. 6.—The charge distribution and field at a sharp p–n junction between homogeneous regions.

The inversion layer has been neglected, which is permissible even with alloyed junctions when a substantial reverse bias is applied.

BC are proportional to the charge densities, and the area of the triangle ABC is proportional to the barrier voltage. The equilibrium barrier voltage or diffusion voltage, V_d, in germanium is generally 200–300 mV. An increase in the reverse bias alters the charge and field distributions to the dotted curves in Fig. 6, where the area of the triangle DEF is proportional to the new value of the applied voltage plus the diffusion voltage, i.e. $V_a + V_d$. The field now stretches further into the n-region.

Fig. 7 represents the field along the axis of a p–n–p transistor with a base of uniform resistivity and no bias on the emitter. Fig. 7(a) shows the approximate field distribution at low voltages. The area of the triangle ABC is proportional to the diffusion

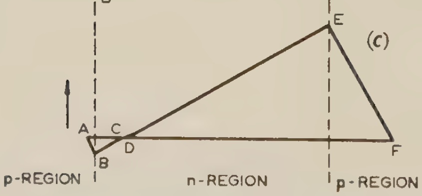


Fig. 7.—The field distribution along the axis of an alloyed p–n–p transistor.

The transistor shows punch-through, the field distribution being shown at successively increasing collector voltages.

voltage of the emitter junction, and the area of the triangle DE corresponds to $V_c + V_d$. The collector depletion layer is th region between D and F, and any holes reaching the point I are swept into the collector in an extremely short time. As th collector voltage rises [Fig. 7(b)] the depletion layer expands, s that any holes emitted into the base have to diffuse only betwee C and D, and a further increase in collector voltage may giv rise to the field distribution shown in Fig. 7(c). This conditio is known as 'punch-through' and is obtainable in a transistc only if the voltage required to stretch the depletion layer all th way across the base is less than the diode breakdown voltag between collector and base, V_B or V_z. If punch-through has bee reached, an increase in collector voltage moves the whole curv BCDE upwards, thereby reducing the area of the triangle ABC This means that the hole concentration at C is substantiall increased, and since the holes have to diffuse only the very shor distance CD, a large current flows. This current is independen of base voltage, provided only that the base is positive or onl slightly negative.

(4.2) The Measurement of Punch-Through Voltage

If a transistor shows punch-through, the voltage at which i occurs (V_p) can be measured in various ways. In general, al will give the same value for V_p, but some methods are mor sensitive than others to any peculiarities in the transistor. Fo instance, if a valve voltmeter is connected between emitter an collector, and the collector-base voltage V_c is increased, th reading on the voltmeter may increase up to a certain voltag and subsequently remain constant. This constant voltage is V_p, but in certain transistors the punch-through is merely at a smal fault in the base between emitter and collector. Generally, th region of the fault has only a very small area and is capable o passing only a small current, owing to the limiting drift velocity of holes and electrons.[9] Therefore, when a transistor with suc a fault is placed in the avalanche circuit it does not behave as i it had the punch-through voltage indicated by this test. The tes may therefore give somewhat misleading results.

An alternative method is to measure the open-circuited emitter turnover voltage, V_B or V_z, and then the turnove voltage with emitter connected to base. If this turnover voltag is substantially lower than V_B or V_z then it is in general equa to V_p. This turnover is shown in Fig. 3. However, a transisto with a high collector leakage current can show in these conditions a turnover which is due not to punch-through but to avalanche multiplication in conjunction with the internal bas resistance (see Section 4.6).

If the base is given a substantial positive bias V_b with respect to the emitter, then a turnover voltage substantially less than $V_B - V_b$ or $V_z - V_b$ can only be due to punch-through. The circuit is shown in Fig. 8. The turnover is generally followed by oscillation, but this need not affect the measurement of V_p which is the voltage (V_{ec}) at which I_c starts to rise steeply.

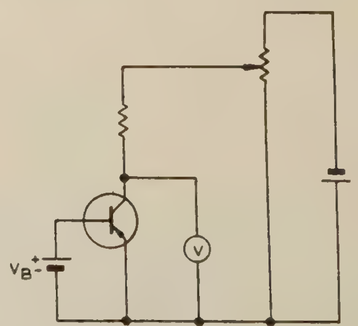


Fig. 8.—The circuit for measuring the punch-through voltage V_p.

(4.3) The Two Types of Avalanche Transistor

The punch-through voltage was measured in the transistors mentioned in Section 1. It was found that the transistors which would oscillate at high speed with no external base resistor and with only stray external collector capacitance (1 or 2 pF) all showed punch-through, V_p being substantially greater than V_x. (This is obviously a necessary condition for oscillation to occur, because otherwise $\alpha_0 M$ can never be appreciably greater than unity.) These will be referred to as punch-through-multiplication (p.t.m.) transistors, and they can readily be distinguished from other avalanche transistors in that their oscillation cannot be controlled by applying a positive voltage to the base. This will be discussed in Section 4.6.

All the p.t.m. transistors gave a very rapid build-up of current. This is to be expected, since at the start of the discharge there is a field right across the base and hence a very short transit time. Thus the current is built up when the transit time is very short, after which V_c falls and the transit time increases, but the current is readily maintained.

The transistors which required both external base resistance and collector capacitance were placed in the circuit shown in Fig. 5. They showed no punch-through, and the oscillation could be controlled by altering the base voltage. Some of these transistors would also operate at very high speeds. It was assumed that this was due to the minimum base width being only just greater than the width of the depletion layer when V_c approached V_B or V_z. The holes would then have to diffuse only a very short way before being swept into the collector by the field. This explanation was confirmed by sectioning some of the transistors.

Another effect which accelerates the build-up of current has been discussed in detail by Schenkel and Statz.[10] The electrons passing into the base from the collector depletion layer increase the emitter-to-base potential. Holes flow into the base from the emitter, and the mutual attraction of the electrons and holes creates a field in the base. The field enables the electrons to move from where they are generated to the region of the base where they are required to neutralize the incoming holes, and assists the holes to diffuse across the base in the opposite direction.

(4.4) The Build-up and Decay of Excess Electrons in the Base.

The other main feature of the relaxation oscillation that requires explanation is the discharge of C_{ex} (Fig. 5) down to almost zero. This is due to the storage of electrons and holes which were produced during the rise of current. The rise of current occurs when the collector reverse current flowing through the base resistance is sufficient to bias the emitter into conduction. Holes flow from the emitter into the collector depletion layer. The collector voltage is substantially greater than V_x, so M is substantially greater than $1/\alpha_0$. Therefore, the electrons produced by the multiplication process are more than enough to supply both the recombination current and the current which flows down the base resistance. The remainder are available for electrostatically neutralizing the incoming holes.

The current which flows down the base resistance causes regenerative feedback and the collector current rises. The collector voltage falls and the depletion layer, even in the p.t.m. transistor, no longer stretches all the way across the base. There is therefore a widening region across which the holes must diffuse, and in this region there must be approximate charge neutrality. The distribution of holes in this region is thus approximately neutralized by an almost similar distribution of the excess electrons.

The shape of this distribution curve of holes and electrons will now be considered. If the base resistance is substantial there can be only a small lateral current flow. Almost all the axial current flow is due to the diffusion of holes, and since the divergence of the total current must be zero, the hole diffusion current must be approximately constant across the base. Thus the concentration gradient of holes across the base is nearly constant. Unless the collector bottoms, the concentration at the edge of the collector depletion layer is effectively zero, and there is therefore an approximately triangular distribution of holes and excess electrons in the base. This distribution causes a current to flow between emitter and collector until the excess electrons have left the base, or the collector bottoms. Therefore, although the generation of electrons effectively ceases when V_c has fallen to V_x, the excess electrons generally take a relatively long time to leave the base, and C_{ex} is usually discharged down to zero. When the electrons have escaped, the collector voltage rises and the cycle can be repeated.

(4.5) Multiplication Factor as a Function of Current and the Minimum Value of C_{ex}

A more detailed study of the ways in which the electrons flow in and out of the base yielded interesting results. The electrons are supplied to the base as a result of the multiplication of the hole current reaching the collector depletion layer. Electrons leave the base by the following means:

(*a*) Recombining with holes (i_r).
(*b*) Flowing through the base resistance (i_b).
(*c*) Flowing into the emitter (i_{ee}) and the collector if it bottoms (i_{ec}).
(*d*) Discharging the internal collector capacitance.

All these ways of removing electrons were considered, but it was the study of the effect of the internal capacitance which led to the most interesting conclusions.

The internal capacitance is due to the widening and narrowing of the depletion layer as the collector voltage rises and falls. When the voltage falls, charge carriers are required to fill up part of the depletion layer, electrons being required on the *n*-side and holes on the *p*-side. Therefore in a *p–n–p* transistor, electrons are absorbed in the base side of the collector depletion layer during the discharge of C_{ex}.

The currents in the transistor are shown in Fig. 9, where the

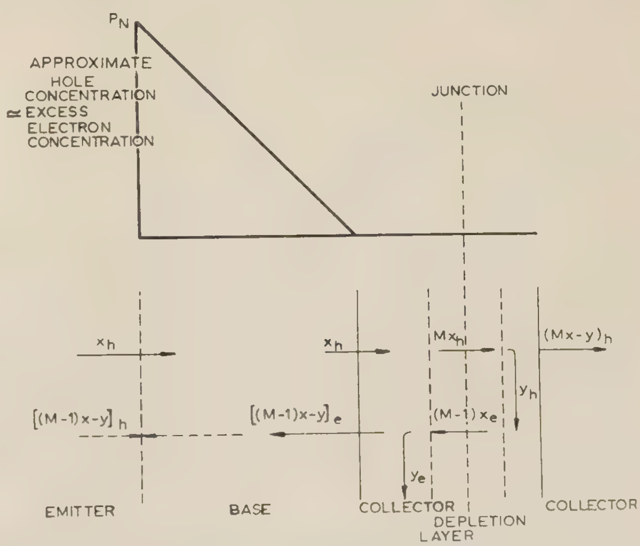


Fig. 9.—Detailed analysis of the effect of the collector depletion layer capacitance (C_c) on the build-up of excess electrons in the base.

hole and electron currents are denoted by suffixes, but in terms of conventional current

$$x_h = -x_e = x$$

and $$y_h = -y_e = y$$

Let x holes pass from the emitter, through the base, and into the collector depletion layer. They are multiplied to Mx in the depletion layer of which y remain in order to discharge the internal capacitor C_c, and $Mx - y$ flow into the collector. Since C_{ex} and C_c have very nearly the same voltage across them (the difference being the forward bias of the emitter), the discharging current of these capacitors must be in the ratio of their capacitances.

Therefore, $$\frac{y}{(Mx - y)} = \frac{C_c}{C_{ex}}$$

i.e. $$y = \frac{Mx}{\frac{C_{ex}}{C_c} + 1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$(M - 1)x$ electrons are generated by multiplication, of which y are absorbed by the depletion layer.

Therefore $(M - 1)x - y$ excess electrons pass into the base.

For the collector voltage to fall to zero the total number of electrons passing into the base during the complete discharge of C_{ex} must be positive.

Therefore $$(M_{eff} - 1)x - y > 0 \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and, substituting eqn. (4) in eqn. (5),

$$M_{eff} > 1 + \frac{C_{c\,eff}}{C_{ex}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where M_{eff} and $C_{c\,eff}$ are the mean values of M and C_c between the value of V_c at the start of the discharge and $V_c = 0$.

If eqn. (1) is used to evaluate M_{eff} and the well-known square-root law[11] to evaluate $C_{c\,eff}$, then a minimum value of C_{ex} for V_c to fall to zero can be found from eqn. (6). In practice it was found that transistors would break down to zero with an external capacitance smaller than the predicted minimum value. This suggested that the value of M was greater than that given by eqn. (1), which was thought to be quite possible owing to the very high current density during the discharge. Eqn. (1) applies for a triangular field distribution, and at high current densities the field distribution is no longer triangular. Owing to the limiting drift velocity of holes and electrons a fairly large number of current carriers are required even in the depletion layer when a very large current has to be carried. The electrostatic charge of the holes that carry the current substantially distorts the field distribution within the collector depletion layer.

Experiments were performed to show whether M did increase at high current densities. As the value of V_B cannot be obtained directly with any accuracy, owing to the likelihood of surface effects, the voltage V_x at which $\alpha_0 M = 1$ was measured at low currents, and the value of V_B calculated from eqn. (1) and the measured value of α_0. Therefore $M - 1$ as a function of voltage could be predicted from eqn. (1). Using short pulses to eliminate heating effects, both V_x and α_0 were measured as functions of current, and the measured value of $M - 1$ was found to be greater than the predicted value by a factor F, as shown in Fig. 10. Since α_0 increases slightly with increasing V_c, owing to depletion-layer widening, the real value of F may be even higher than that shown in Fig. 10.

There is a very wide variation in F between different transistors owing to their different geometry. The distribution of current is strongly dependent on the exact geometry, especially near punch-through when the current is highly localized at the point or points at which the collector depletion layer is nearest the emitter. At or near punch-through very high local current densities can be expected even with quite moderate currents. Quite large values of F are therefore probably obtained, particularly in the p.t.m. transistor, and the mean value of M is substantially greater than that predicted by eqn. (1). This is a partial explanation of why the p.t.m. transistors require so low an external capacitance.

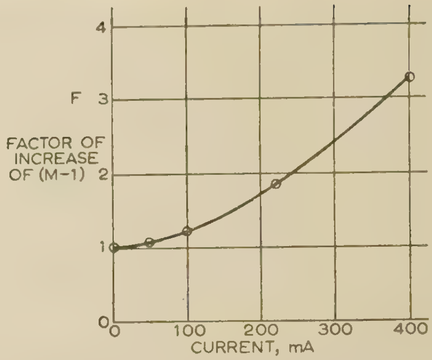


Fig. 10.—The increase of M with current in an alloyed *p–n–p* germanium transistor.

The value of F varies greatly between different transistors.

It is worth commenting on the other ways in which electrons leave the base. The following results have been obtained from well-known transistor theory:[12]

(*a*) i_r is approximately proportional to the hole concentration in that part of the base nearest the emitter, p_N. (Equilibrium hole concentration is p_{N0}.)

(*b*) i_{ec} is negligible unless the transistor bottoms, in which case it is of the same order of magnitude as i_{ee}.

(*c*) i_{ee} is proportional to p_N^2 at the high current densities in these transistors.

(*d*) i_b is less dependent on p_N than $\log(p_N/p_{N0})$, and is in general substantially independent of p_N.

(*e*) p_N is roughly proportional to C_{ex}.

Thus, as C_{ex} increases, i_{ee} becomes relatively more important and i_b less important.

(4.6) Analysis of the Start of the Discharge

The build-up of current has been qualitatively explained as being due to regenerative feedback, but a semi-quantitative understanding of the process can be obtained from Fig. 11. The current distribution is unstable when R_{in} is negative.

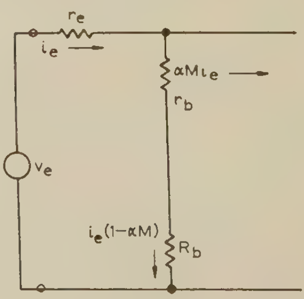


Fig. 11.—The equivalent input circuit of a transistor at low frequencies.

$$v_e \doteq i_e r_e + i_e(1 - \alpha_0 M)(R_b + r_b)$$

$$R_{in} = \frac{v_e}{i_e} = r_e + (1 - \alpha_0 M)(R_b + r_b)$$

Instability therefore occurs when

$$r_e < (\alpha_0 M - 1)(R_b + r_b) \quad . \quad . \quad . \quad . \quad (7)$$

When a transistor is operating normally, r_e can be determined from the equation

$$I_e = I_k \exp (qV_{eb}/kT) \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where V_{EB} is the emitter-base voltage, and I_k is a constant. Differentiating eqn. (8) gives

$$r_e = dV_{eb}/dI_e = 0{\cdot}026/I_e \quad . \quad . \quad . \quad . \quad (9)$$

In the p.t.m. transistor this relationship is only very approximate because I_k is not really a constant owing to the very large Early effect.[13]

From eqns. (7) and (9), instability occurs approximately when

$$I_e > \frac{0{\cdot}026}{(\alpha_0 M - 1)(R_b + r_b)} \quad . \quad . \quad . \quad . \quad (10)$$

In the p.t.m. transistor the internal base resistance increases rapidly as the collector voltage approaches V_p. The reason is that the collector depletion layer, being depleted of mobile carriers, contributes almost nothing to the lateral conductance. Thus the area over which there is lateral current flow is reduced as the depletion layer expands, until at punch-through the lateral resistance is very high indeed. This high lateral resistance corresponds to a high base resistance, and the electron current from the collector depletion layer which flows into the punch-through region of the base develops a very high voltage between this region and the base lead of the transistor. This is why the base voltage has so little effect on the operation of the p.t.m. transistor. The high base resistance also causes instability at very low emitter currents, as can be seen from eqn. (10).

(5) THE DESIGN OF P.T.M. AND HIGH-SPEED AVALANCHE TRANSISTORS

Various transistors have been made to test the theories that have been discussed. Since the base must be of a fairly high resistivity to take advantage of the width of the depletion layer, the transistors were made by the alloying technique. The most important equation for the design of these transistors is

$$w_d = K\sqrt{(\rho V_c)}$$

where w_d = Depletion-layer width, microns.
ρ = Resistivity of the base, ohm-cm.
V_c = Collector voltage.

For germanium *p–n–p* transistors $K \simeq 1$. For p.t.m. transistors the alloying conditions must be such that when V_c is somewhere between V_B and V_x, w_d is equal to the minimum base width. For normal high-speed avalanche transistors the minimum base width should be just greater than the value of w_d when $V_c = V_B$.

The other properties of the transistor which must be closely controlled are V_B and the collector capacitance. V_B is determined by the resistivity of the base, and in the transistors used in the circuits described in Sections 7 and 8 V_B was about 55 volts. The internal collector capacitance was kept fairly low by using 400 μ diameter indium-gallium pellets, and it was found that an external capacitance of 30 pF was enough to cause the collector to bottom [see eqn. (6)]. The transistors were symmetrical, since the normal α' was found to be of almost no consequence, and the better junction could then be chosen as the collector.

(6) APPLICATIONS

Both the p.t.m. and the high-speed avalanche transistor can be used as negative-resistance devices. Theoretically this could lead to their application in pulse circuits and line communication circuits, but, as it is difficult with present transistors to provide a stable known value of negative resistance, applications are likely to be confined to pulse circuits.

(6.1) Application of the P.T.M. Transistor

As mentioned in Section 1, the p.t.m. transistor behaves like a 2-terminal negative resistance, and its application to pulse circuits can be compared with that of the cold-cathode gas-filled triode. The punch-through voltage is equivalent to the striking voltage, and the voltage at which $\alpha_0 M = 1$ is equivalent to the burning voltage. The main difference between the two is that the trigger electrode in the cold-cathode gas-filled triode is operative before striking and then loses control, whereas the base electrode of the p.t.m. transistor has no control at punch-through and regains control in the region where $\alpha_0 M = 1$. The other differences are obvious; the p.t.m. transistor operates at a lower voltage and at a much higher speed.

The main difficulty with the p.t.m. transistor arises from the very small external capacitance required to produce relaxation oscillation. It is therefore very difficult to produce a circuit with two stable states, and its main use is as a relaxation oscillator for providing very fast edges and very short current pulses.

(6.2) The Application of the High-Speed Avalanche Transistor

As the avalanche transistor has a current gain greater than unity it may be used in point-contact transistor circuits. The bottomed voltage of the point-contact transistor is equivalent to the voltage at which $\alpha_0 M = 1$ in the avalanche transistor. This means that bistable circuits using the avalanche transistor are subject to a high collector dissipation in the 'on' state.

The avalanche transistor may also be used as a relaxation oscillator in a similar manner to the p.t.m. transistor, with the addition of an external capacitor and an external base resistor. However, the external capacitance required for the high-speed avalanche transistor is small, and high-speed high-repetition-rate pulse generators can be made which can be controlled by the base electrode.

(7) CIRCUIT INVESTIGATIONS

From the brief survey of applications outlined above it appeared that the most promising line of investigation lay in the operation of the p.t.m. and high-speed avalanche transistors in the relaxation mode.

(7.1) The P.T.M. Transistor

(7.1.1) Circuit Operation.

On applying a voltage $-V_{cc}$ to the circuit of Fig. 12 using a p.t.m. transistor, the capacitor C charges through R_c towards the

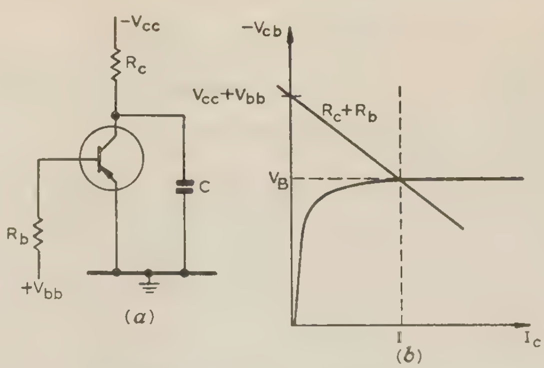


Fig. 12.—Avalanche transistor.
(a) Basic circuit of relaxation oscillator.
(b) Static characteristic and load line.

value $-V_{cc}$. When the voltage at the collector reaches V_p the transistor conducts and the capacitor discharges. As soon as conduction ceases, the capacitor starts to recharge, and the cycle recommences.

An output voltage may be taken in two ways: (a) from across the capacitor or a portion of a capacitance potentiometer, or (b) from across a small resistor or transformer in series with the discharge path.

The only way in which the oscillation can be controlled at the base is by making the base voltage greater than $V_B - V_p$. In this way the collector voltage may be prevented from rising to V_p. However, this is not a very satisfactory method of control for a triggered circuit, since it is difficult to make transistors with closely controlled values of V_p and V_B, and there would therefore be a large variation in trigger sensitivity.

(7.1.2) **Transistor Dissipation and Maximum Repetition Rate.**

In calculating the dissipation in the transistor when it is working as a relaxation oscillator with an external capacitor, there are three factors to be considered:

(a) The capacitor discharge current.
(b) Additional collector current through R_c.
(c) The collector voltage and current during the recharging of the capacitor.

The discharge of the capacitor provides a total energy of $\frac{1}{2}CV_p^2$. This must be averaged over the complete cycle, although there will also be a maximum permissible peak dissipation for any given transistor.

Additional collector current through R_c occurs only during the discharge period, and with normal values of load resistor is of at least an order of magnitude lower than the capacitor current.

During the recharging of the capacitor the collector current is negligible.

Thus the mean dissipation in the transistor is given by

$$P = \frac{CV_p^2}{2T} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where T is the time of one cycle of operation and the effect of the internal collector capacitance has been neglected.

Thus the maximum repetition rate is limited by the dissipation in the transistor.

Since the discharge time is a negligible fraction of the total period, the time taken to recharge is approximately equal to T:

$$V_p = V_{cc}(1 - \varepsilon^{-T/CR_c})$$

so that
$$T = CR_c \log_\varepsilon \frac{V_{cc}}{V_{cc} - V_p} \quad . \quad . \quad . \quad . \quad (12)$$

Substituting for T in eqn. (11),

$$P = \frac{V_p^2}{2R_c} \frac{1}{\log_\varepsilon [V_{cc}/(V_{cc} - V_p)]} \quad . \quad . \quad . \quad (13)$$

(a) Eqn. (13) shows that the dissipation is independent of the value of C.
(b) For a given value of dissipation eqn. (13) gives a fixed relationship between R_c and V_{cc}.
(c) V_{cc} may be fixed arbitrarily, but, in order that $V_{cc}/(V_{cc} - V_p)$ should vary as little as possible with variations of V_p in different transistors, V_{cc} should be made as large as possible.
(d) Having fixed V_{cc}, the value of R_c can be calculated from eqn. (13).
(e) The value of C can now be chosen to suit the repetition rate and the value of peak current required.

In cases where the external capacitance is small compared with the internal collector capacitance, the cycle time depends mainly on the recharging of the internal capacitance through the base resistance and the collector load. The current through the base resistance may prevent the transistor cutting off during the initia part of the recharging cycle, but this appears to have little effect on the operation.

(7.2) **The High-Speed Avalanche Transistor**

(7.2.1) **Circuit Operation.**

On applying a voltage $-V_{cc}$ to the circuit of Fig. 12 the capacitor C charges towards the value $-V_{cc}$ through the load resistor R_c. The collector-base current of the transistor therefore rises towards I. The value of R_b is chosen so that $R_bI > V_{bb}$. Thus the emitter is brought into conduction, and current multiplication in the transistor causes the capacitor C to discharge completely in a short time. After the discharge, the collector-base current decreases, and the transistor is cut off by the positive voltage applied to the base.

If R_b is such that $R_bI < V_{bb}$, then the circuit will have a stable state in which the current I flows from base to collector, and the emitter current is zero. If the base is now made negative with respect to the emitter by means of an applied trigger pulse, the emitter conducts and regeneration causes the capacitor C to discharge through the transistor. The minimum amplitude of trigger pulse required is approximately $V_{bb} - IR_b$. Since V_B and hence I may vary from one transistor to another, it is preferable that R_b should be low in value. The minimum possible value is that of the internal base resistance of the transistor.

The trigger pulse may be applied as a negative pulse to the base or as a positive pulse to the emitter. If the trigger is applied to the base R_b will act as a shunt, so that more trigger power will be required when R_b has a low value.

With a resistor in the emitter circuit, the same conditions apply, since the resistor is in the capacitor discharge path and must therefore be of low value. However, since the emitter is driven positive to trigger, and the discharge drives it negative, a diode may be used in the emitter circuit. This presents a high shunt resistance to the trigger pulse and a low series resistance to the discharge current.

With this circuit arrangement there is no longer the necessity for external base resistance, and the value of V_{bb} may be made small, thereby stabilizing the trigger sensitivity at a high value. A value of V_{bb} of approximately 0·3 volt may be achieved by

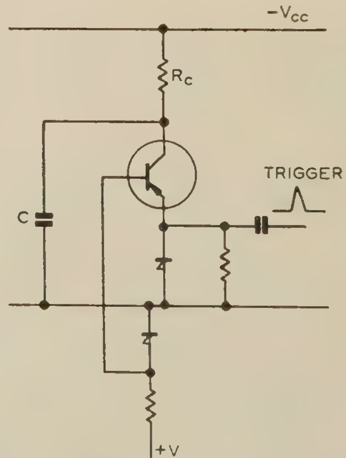


Fig. 13.—The circuit of a regenerative pulse amplifier using the high-speed avalanche transistor.

biasing a diode in the forward direction, as shown in Fig. 13. An added advantage of this method of bias is that during the discharge time sufficient current flows in the base to reverse the bias on the diode, so that the switching is not affected by a

constant positive bias on the base tending to hold the transistor off.

(7.2.2) Transistor Dissipation and the Maximum Repetition Rate.

With the p.t.m. transistor, only the operating dissipation had to be considered: with the high-speed avalanche transistor in a triggered circuit there may be a quiescent state. In certain conditions the quiescent dissipation will exceed the operating dissipation and it is therefore necessary to determine the relationship between them.

From Fig. 12(*b*), I is given by

$$I = \frac{V_{cc} - V_B}{R_c}$$

so that the dissipation in the quiescent state is

$$P_q = \frac{V_B(V_{cc} - V_B)}{R_c} \quad . \quad . \quad . \quad . \quad (14)$$

The dissipation during the operating cycle, P_o, is the same as for the p.t.m. transistor except that the maximum collector voltage is V_B instead of V_p.

Eqns. (11) and (13) now become

$$P_o = \frac{CV_B^2}{2T} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

$$P_o = \frac{V_B^2}{2R_c} \frac{1}{\log_\varepsilon [V_{cc}/(V_{cc} - V_B)]} \quad . \quad . \quad . \quad (16)$$

From eqns. (14) and (16),

$$\frac{P_q}{P_o} = 2(s - 1) \log_\varepsilon \frac{s}{s - 1} \quad . \quad . \quad . \quad . \quad (17)$$

where $s = V_{cc}/V_B$.

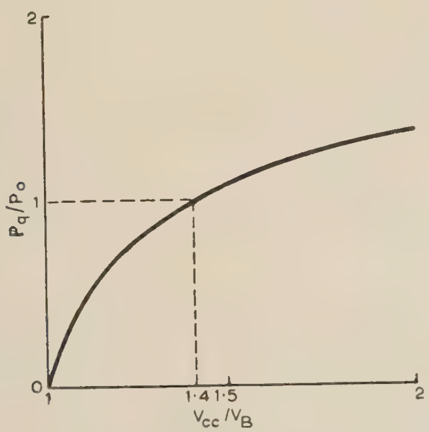


Fig. 14.—The ratio of the quiescent and operating dissipations as a function of supply voltage.

The ratio P_q/P_o is plotted against V_{cc}/V_B in Fig. 14. From this it can be seen that below a certain value of V_{cc}/V_B, $P_q < P_o$.

Spreads in V_B have to be considered in the same way as the spreads in V_p with the p.t.m. transistor. V_{cc} will therefore have to be fixed as a compromise between the highest possible value to reduce the effect of spreads and the lowest possible value to reduce the quiescent dissipation.

Two design procedures then follow:

(i) If the value of V_{cc} is such that $P_q > P_o$ the value of R_c is calculated from eqn. (14).
(ii) If the value of V_{cc} is such that $P_q < P_o$ the value of R_c is calculated from eqn. (16).

In both cases C may then be chosen to suit the repetition rate and the value of peak current required.

If the mark/space ratio is fixed and $P_q < P_o$, then P_o may be increased provided that the allowable peak dissipation of the transistor is not exceeded.

Where it is desired to operate a circuit at any repetition rate up to the maximum possible, the operating dissipation is made equal to the maximum permissible dissipation of the transistor. The value of V_{cc} must then be chosen to make the quiescent dissipation equal to or less than the operating dissipation.

The above calculations are based on an ideal characteristic. In practice it will not be possible to operate the transistor at the avalanche voltage V_B owing to curvature of the turnover characteristic. The transistor will operate at a slightly lower voltage and the real dissipation will be less than the calculated one.

(7.2.3) Design of Practical Circuits.

The transistors designed to operate in the regenerative amplifier circuit had the following parameters:

$$V_B = 65 \text{ volts}; \; P_{max} = 100\,\text{mW}$$

and the minimum external capacitance required for complete breakdown was less than 30 pF.

Two examples of circuits using these transistors will be given.

Example I.—A regenerative amplifier to operate at a repetition rate of 1 Mc/s.

V_c is fixed as a compromise at the value giving $P_o = P_q$.

Therefore $\qquad V_{cc} = 1{\cdot}4 \times 65 \simeq 90$ volts

From eqn. (16),

$$R_c = \frac{65^2}{0{\cdot}2} \frac{1}{\log (1{\cdot}4/0{\cdot}4)} = 16\,650 \text{ ohms}$$

For the circuit to be capable of operating at a repetition rate of 1 Mc/s, T must be less than 1 microsec. If we make $T = 0{\cdot}8$ microsec, then, from eqn. (15),

$$C = \frac{0{\cdot}1 \times 2 \times 0{\cdot}8 \times 10^{-8}}{65^2} = 38\,\text{pF}$$

Example II.—A pulse generator to provide current pulses with a peak value of 1 amp into a low-impedance load.

The current pulse is approximately of the form of a half-sine-wave and is composed of the total charge on the capacitor flowing out during the discharge time t, which is typically 0·05 microsec.

Hence $\qquad \frac{2}{\pi} I_{pk} t = CV_B$

so that $C = 490$ pF for $I_{pk} = 1$ amp.

As before, with $V_{cc}/V_B = 1{\cdot}4$,

$$T_{min} = \frac{\frac{1}{2}CV_B^2}{P_{max}} = 10{\cdot}3 \text{ microsec}$$

so that the maximum repetition rate is approximately 100 kc/s.

Circuits based on these designs have been constructed and operate substantially as calculated. A practical circuit based on Example I is shown in Fig. 15, together with the voltage and current waveforms obtained. The performance of the circuit is unaffected by changes in ambient temperature from 20° C to 45° C. This regenerative amplifier has been used in conjunction with a delay line to form a circulating store operating at a repetition frequency of 1 Mc/s.

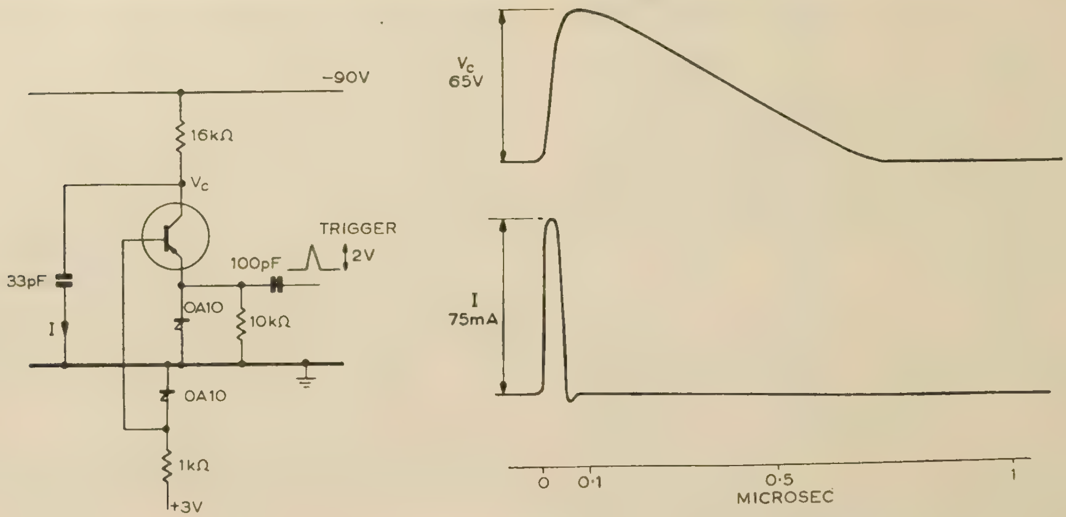


Fig. 15.—Practical regenerative-amplifier circuit, with waveforms.

(8) CONCLUSIONS

The paper has described techniques which will produce very fast pulse generators using the simple alloy transistor. The p.t.m. and high-speed avalanche transistor have been described in some detail and should prove useful in many applications if the manufacturing problems can be overcome. The problems arise mainly from the necessity of accurately controlling the minimum base width, as shown in Section 5. A possible approach would be to select suitable transistors from production r.f. alloyed transistors.

The operating voltage decreases with decreasing resistivity of the base, but this also decreases the width of the depletion layer. This means that the minimum base width would have to be even more closely controlled if low operating voltage were required. In addition, the results of Knott *et al.*[3] indicate that at resistivities below a certain value, M can never be great enough to cause regeneration, owing to Zener breakdown. However, this conclusion may not be correct because of the apparent increase in M at high current densities.

The ionization process in semiconductors has proved to be extremely fast. The performance of the p.t.m. and high-speed avalanche transistors is equivalent to that of normal transistors with a cut-off frequency more than an order of magnitude greater. The experimental transistors described have a performance equivalent to transistors with a cut-off frequency of 100 Mc/s operating in conventional circuits. Further investigation into the application of this process may produce a new range of devices which will be the semi-conductor equivalents of the present range of gas-discharge tubes but operating at much higher speeds.

(9) ACKNOWLEDGMENTS

The authors wish to acknowledge the valuable assistance of their colleague Mr. T. B. Watkins.

(10) REFERENCES

(1) ZENER, C.: 'Theory of the Electrical Breakdown of Solid Dielectrics', *Proceedings of the Royal Society*, A, 1934, **145**, p. 523.

(2) MCKAY, K. G., and MCAFFEE, K. B.: 'Electron Multiplication in Silicon and Germanium', *Physical Review*, 1954, **61**, p. 1079.

(3) KNOTT, R. D., COLSON, I. D., and YOUNG, M. R. P.: 'Breakdown Effect in *p–n* Alloy Germanium Junctions', *Proceedings of the Physical Society*, 1955, **65**, p. 182.

(4) WOLFENDALE, E.: 'Avalanche Breakdown in Junction Transistors', British Patent Application No. 18498: 1955.

(5) MILLER, S. R., and EBERS, J. J.: 'Alloyed Junction Avalanche Transistors', *Bell System Technical Journal*, 1955, **34**, p. 883.

(6) KIDD, M. C., HASENBERG, W., and WEBSTER, W. M.: 'Delayed Collector Conduction, a New Effect in Junction Transistors', *R.C.A. Review*, 1955, **16**, p. 16.

(7) PRITCHARD, R. L.: 'Frequency Variation of Current Amplification Factor for Junction Transistors', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1476.

(8) SCHENKEL, H., and STATZ, H.: 'Voltage Punch-Through and Avalanche Breakdown and their Effect on the Maximum Operating Voltages for Junction Transistors', *Proceedings of the National Electronics Conference*, 1954, **10**, p. 614.

(9) SHOCKLEY, W.: 'Hot Electrons in Germanium and Ohm's Law', *Bell System Technical Journal*, 1951, **30**, p. 990.

(10) SCHENKEL, H., and STATZ, H.: 'Junction Transistors with Alpha greater than Unity', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 360.

(11) SHOCKLEY, W.: 'The Theory of *p–n* Junctions in Semi-Conductors and *p–n* Junction Transistors', *Bell System Technical Journal*, 1949, **28**, p. 450.

(12) WEBSTER, W. M.: 'On the Variation of Junction-Transistor Current Amplification Factor with Emitter Current', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 914.

(13) EARLY, J. M.: 'Effects of Space-Charge Layer Widening in Junction Transistors', *ibid.*, 1952, **40**, p. 1401.

[6]21.372.823.2

Paper No. 2369 R
July 1957

PROPAGATION OF THE CIRCULAR H_{01} LOW-LOSS WAVE MODE AROUND BENDS IN TUBULAR METAL WAVEGUIDE

By Prof. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), Member.

(*The paper was first received* 16*th January, and in revised form* 11*th March*, 1957.)

SUMMARY

Although there is ample evidence of the suitability of a straight length of hollow metal tube as a waveguide supporting the circular H_{01}-mode for long-distance communication, bends in the guide still present a serious problem. An examination of the conditions required to maintain as nearly as possible the appropriate field distribution at a bend leads to the important conclusion that the wavefront, represented by an equi-phase plane, must remain radial with respect to the centre of curvature. The evidence shows that there are two ways of approaching this requirement: (*a*) by varying the permittivity of the dielectric medium inside the guide over its cross-section, and (*b*) by varying the surface reactance of the guide around its circumference. The first of these proposals is simpler in application and more likely to meet the need, if suitable cellular polystyrene or polythene can be obtained for the purpose.

(1) INTRODUCTION

The problem of maintaining the support of the circular H_{01} wave mode at bends in a tubular metal waveguide has been the subject of much discussion and experiment in recent years, but a completely satisfactory solution has not yet apparently been found. The great interest in this wave mode arises from its unique low-attenuation characteristic when propagated along the inside of a hollow metal tube at a frequency well above the cut-off value. The practicability of the arrangement when the guide is straight has been well established,[1] and at a frequency of 5000 Mc/s attenuations of the order of 14 dB per mile in a copper tube of 1·72 cm inside radius ($1\frac{1}{2}$ in outside diameter), can be readily obtained. This performance takes into consideration variations in cross-section of the tube normally to be expected in commercial production and also tolerances in longitudinal alignment which are not difficult to meet. However, in practical applications of such a guide, e.g. for a long-distance communication channel, deliberate bends are inevitable, and consequently it is of vital importance to know how to negotiate such bends without introducing serious additional losses.

Jouguet[2] showed theoretically that, for gradual bends in guides of a given radius r_0 operated at a free-space wavelength λ_0, there was a favourable angle, θ_h, subtended at the centre of curvature for which the output was substantially a replica of the H_{01} wave input. The relationship derived by Jouguet is

$$\theta_h = \frac{2\pi n}{2\cdot 32}\left(\frac{\lambda_0}{r_0}\right)$$

where n is an integer.

This was subsequently confirmed experimentally by Sims.[3] Whilst this provides a useful means of changing the direction of the waveguide through definite angles related to the size of the tube employed and the wavelength of operation, it does not give the flexibility required in practice. Miller[4] and his colleagues have applied circularly corrugated guides of various kinds, primarily with the object of breaking the degeneracy between the H_{01}-mode and the E_{11}-mode, which tends to be excited at bends. In this way it is claimed that coupling between the wanted and unwanted modes over a length of guide is reduced, and if lossy material is introduced into some of the corrugations, spurious energy can be absorbed. Extensive experiment has shown that, whilst this device is helpful, it does not completely solve the problem and further steps are necessary. Recently a new proposal has been put forward[5] using a hollow tube made of thin copper wire very tightly coiled and wrapped inside a flexible outer coating which holds the coiled wire in place. No details of this development have so far been given, and no figures are available about the radius of curvature that can be tolerated. Published pictorial information suggests that comparatively sharp bends of a few yards' radius cannot be dealt with in this way.

(2) THEORETICAL CONSIDERATIONS

(2.1) Basic Field Equations

To examine the problem of wave propagation in a curved tubular waveguide it is convenient to set up the field equations in toroidal co-ordinates (Fig. 1). For curl E we then find:

$$\frac{1}{r}\left(E_\phi + r\frac{\partial E_\phi}{\partial r} - \frac{\partial E_r}{\partial \phi}\right) = -\mu\frac{\partial H_\theta}{\partial t} \quad . \quad . \quad . \quad (1)$$

$$\frac{1}{R}\left(\frac{\partial E_r}{\partial \theta} - R\frac{\partial E_\theta}{\partial r} - E_\theta \cos\phi\right) = -\mu\frac{\partial H_\phi}{\partial t} \quad . \quad . \quad (2)$$

$$\frac{1}{r}\frac{\partial E_\theta}{\partial \phi} - \frac{1}{R}\frac{\partial E_\phi}{\partial \theta} - \frac{1}{R}E_\theta \sin\phi = -\mu\frac{\partial H_r}{\partial t} \quad . \quad . \quad (3)$$

and similarly for curl H we get:

$$\frac{1}{r}\left(H_\phi + r\frac{\partial H_\phi}{\partial r} - \frac{\partial H_r}{\partial \phi}\right) = \sigma E_\theta + \epsilon\frac{\partial E_\theta}{\partial t} \quad . \quad . \quad (4)$$

$$\frac{1}{R}\left(\frac{\partial H_r}{\partial \theta} - R\frac{\partial H_\theta}{\partial r} - H_\theta \cos\phi\right) = \sigma E_\phi + \epsilon\frac{\partial E_\phi}{\partial t} \quad . \quad (5)$$

$$\frac{1}{r}\frac{\partial H_\theta}{\partial \phi} - \frac{1}{R}\frac{\partial H_\phi}{\partial \theta} - \frac{1}{R}H_\theta \sin\phi = \sigma E_r + \epsilon\frac{\partial E_r}{\partial t} \quad . \quad (6)$$

where $$R = R_a + r\cos\phi \quad . \quad . \quad . \quad . \quad . \quad (7)$$

(2.2) Wave Equation for Low-Loss Circular H_{01}-mode

The low-loss feature of the circular H_{01}-mode arises from the absence of longitudinal current in the wall of the guide, and the only field components appropriate to this mode are E_ϕ, H_r and H_θ. If, in the first place, we assume that the curved guide is capable of supporting such a mode and that we have a sinusoidal time variation, eqns. (1)–(6) reduce to

$$\frac{1}{r}\left(\mathscr{E}_\phi + r\frac{\partial \mathscr{E}_\phi}{\partial r}\right) = -j\omega\mu\mathscr{H}_\theta \quad . \quad . \quad . \quad . \quad (8)$$

$$-\frac{1}{R}\frac{\partial \mathscr{E}_\phi}{\partial \theta} = -j\omega\mu\mathscr{H}_r \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Written contributions on papers published without being read at meetings are [i]nvited for consideration with a view to publication.
Prof. Barlow is Pender Professor of Electrical Engineering, University College, [Lon]don.

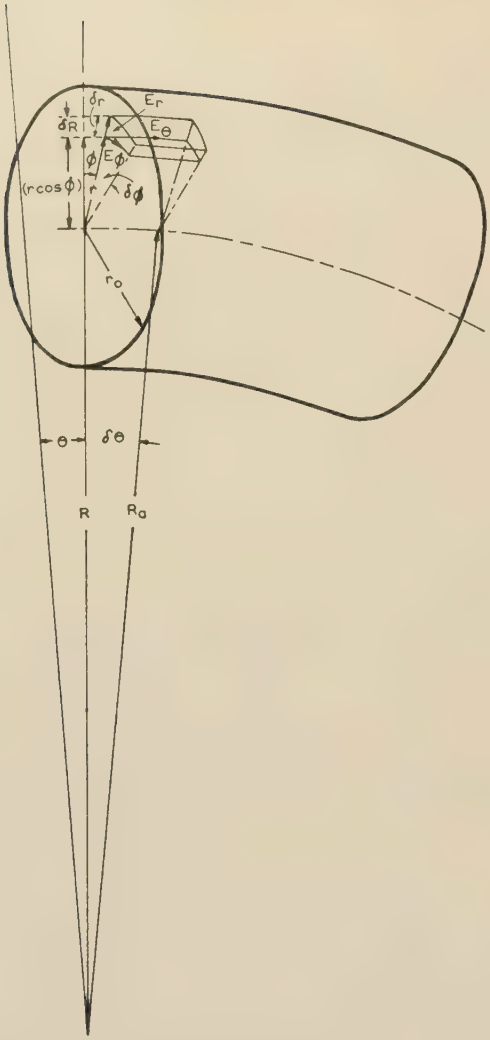


Fig. 1.—Toroidal co-ordinates for curved waveguide.

$$-\frac{1}{r}\frac{\partial \mathscr{H}_r}{\partial \phi} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\frac{1}{R}\frac{\partial \mathscr{H}_r}{\partial \theta} - \frac{\partial \mathscr{H}_\theta}{\partial r} - \frac{1}{R}\mathscr{H}_\theta \cos\phi = (\sigma + j\omega\epsilon)\mathscr{E}_\phi \quad . \quad (11)$$

$$\frac{1}{r}\frac{\partial \mathscr{H}_\theta}{\partial \phi} - \frac{1}{R}\mathscr{H}_\theta \sin\phi = 0 \quad . \quad . \quad . \quad . \quad (12)$$

where $\mathscr{E}_\phi$, $\mathscr{H}_r$ and $\mathscr{H}_\theta$ are the field components as functions of the space co-ordinates only. From eqns. (8), (9) and (11) we therefore get

$$\frac{1}{R^2}\frac{\partial^2 \mathscr{E}_\phi}{\partial \theta^2} + \frac{\partial^2 \mathscr{E}_\phi}{\partial r^2} + \left(\frac{1}{r} + \frac{\cos\phi}{R}\right)\frac{\partial \mathscr{E}_\phi}{\partial r} + \left(\frac{\cos\phi}{rR} - \frac{1}{r^2}\right)\mathscr{E}_\phi = P^2 \mathscr{E}_\phi \quad . \quad . \quad . \quad . \quad (13)$$

where $\qquad P^2 = j\omega\mu(\sigma + j\omega\epsilon)$

and assuming a perfect dielectric medium through which the wave is propagated,

$$\sigma = 0$$

so that $\qquad P^2 = -\,\omega^2\mu\epsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$

The variation of $\mathscr{E}_\phi$ with θ is represented by

$$\mathscr{E}_\phi = A_1 \mathscr{E}_{\phi r} \varepsilon^{-\gamma R \theta} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where $\mathscr{E}_{\phi r}$ is that part of $\mathscr{E}_\phi$ which is dependent upon r and A_1 is constant.

The longitudinal propagation coefficient γ is $\alpha + j\beta$ and neglecting losses, $\alpha = 0$, so that

$$\gamma = j\beta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

The phase-change coefficient β is normally regarded as a constant, but this leads to a wave equation which apparently cannot yield, on any reasonable approximation, a solution representing the desired mode.

(*a*) When $\qquad \beta = \text{constant} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$

we have

$$\frac{\partial^2 \mathscr{E}_{\phi r}}{\partial r^2} + \left(\frac{1}{r} + \frac{\cos\phi}{R} - 2\gamma\theta\cos\phi\right)\frac{\partial \mathscr{E}_{\phi r}}{\partial r} + \left[\gamma^2 - P^2 + \left(\frac{\cos\phi}{rR} - \frac{1}{r^2}\right) - \left(\frac{1}{r} + \frac{\cos\phi}{R}\right)\gamma\theta\cos\phi + (\gamma\theta\cos\phi)^2\right]\mathscr{E}_{\phi r} = 0 \quad (18)$$

Both on the outside of the bend when $\phi = 0$ and on the inside of the bend when $\phi = 180°$ the term $\gamma\theta\cos\phi$ in eqn. (18) becomes very significant, and it seems impossible to get a near approach to the Bessel equation which leads to the circular H_{01} mode.

(*b*) When $\qquad \beta R = \text{constant} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$

eqns. (13) and (15) give

$$\frac{\partial^2 \mathscr{E}_{\phi r}}{\partial r^2} + \left(\frac{1}{r} + \frac{\cos\phi}{R}\right)\frac{\partial \mathscr{E}_{\phi r}}{\partial r} + \left(\gamma^2 - P^2 - \frac{1}{r^2} + \frac{\cos\phi}{rR}\right)\mathscr{E}_{\phi r} = 0 \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Since $R \gg r$ in any practical application of this guide, we can neglect $(\cos\phi)/R$ in comparison with $1/r$, and eqn. (20) can be rewritten, as a good approximation,

$$\frac{\partial^2 \mathscr{E}_{\phi r}}{\partial r^2} + \frac{1}{r}\frac{\partial \mathscr{E}_{\phi r}}{\partial r} + \left(U^2 - \frac{1}{r^2}\right)\mathscr{E}_{\phi r} = 0 \quad . \quad . \quad (21)$$

where $\qquad U^2 = \gamma^2 - P^2 = \omega^2\mu\epsilon - \beta^2 \quad . \quad . \quad . \quad . \quad (22)$

Condition (19) means that the wavefront, as represented by an equi-phase plane, is required to be radial with respect to the centre of curvature of the bend in the guide. This would appear to be a vital condition for the maintenance of a near approach to the low-loss mode at a bend.

If $\beta = \beta_a$ when $R = R_a$ (i.e. $\phi = 90°$ or $270°$), we must have $\beta R = \beta_a R_a$, and using eqn. (7), we get

$$\beta = \frac{\beta_a}{1 + \dfrac{r}{R_a}\cos\phi} \quad . \quad . \quad . \quad . \quad (23)$$

It must be emphasized that eqn. (20) has been obtained on the assumption that only three components of field, namely $\mathscr{E}_\phi$, $\mathscr{H}_r$ and $\mathscr{H}_\theta$, are present in the guide. Examination of eqns. (8)–(12) shows that they are not entirely consistent and that this assumption cannot therefore be strictly true although the error to which it leads is small when βR is constant and $R \gg r$ (see Section 7).

(2.3) Approximate Solutions to Wave Equation and Conditions Imposed

(2.3.1) Variation of the Permittivity of the Dielectric over the Cross-Section of the Guide when Bent.

In order to solve eqn. (21) as the standard form of Bessel's equation, we require that U^2 should be a constant. This is not consistent with eqns. (22) and (23) if, as is usual, μ and ϵ for the medium inside the guide are constants. The value of μ cannot readily be varied with r and ϕ, but it is possible to meet the requirements by changing the value of ϵ. Thus suppose $\epsilon = \epsilon_a$ when $R = R_a$ and $\beta = \beta_a$. From eqn. (22),

$$U^2 = \omega^2\mu\epsilon - \beta^2 = \omega^2\mu\epsilon_a - \beta_a^2 = \text{constant}$$

and $$\beta^2 - \beta_a^2 = \omega^2\mu(\epsilon - \epsilon_a)$$

Hence we can write, as a good approximation, since $R \gg r$,

$$\epsilon = \epsilon_a - \frac{2\beta_a^2 r\cos\phi}{\omega^2\mu R_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

According to eqn. (24) the curved guide requires only a comparatively small change over the cross-section in the permittivity of the dielectric, and the highest value is on the inside of the bend.

Assuming that provision is made for this change of permittivity so that U^2 remains constant, we find, from eqn. (21),

$$\mathscr{E}_{\phi_r} = A_2\mathrm{J}_1(Ur) + A_3\mathrm{Y}_1(Ur)$$

and since $\mathscr{E}_{\phi_r} = 0$ when $r = 0$ we get the usual form of equation for the electric field of the circular H_{01}-mode:

$$\mathscr{E}_\phi = A\varepsilon^{-j\beta R\theta}\mathrm{J}_1(Ur) \quad . \quad . \quad . \quad . \quad (25)$$

From eqn. (8) we also have, for the longitudinal magnetic field,

$$\mathscr{H}_\theta = -\frac{UA}{j\omega\mu}\varepsilon^{-j\beta R\theta}\mathrm{J}_0(Ur) \quad . \quad . \quad . \quad . \quad (26)$$

where $$U = \sqrt{(\omega^2\mu\epsilon_a - \beta_a^2)} \quad . \quad . \quad . \quad . \quad (27)$$

Provided that practical arrangements can be made to bring about the required change of permittivity ϵ over the cross-section of the guide when it is bent, this proposal appears to offer an attractive solution to the problem. Theoretically we should expect to be able to maintain in this way a close approach to the required field distribution, even when the guide is bent with a radius of curvature as small as 50 to 100 times the radius of the tube.

(2.3.2) Variation of the Surface Reactance around the Periphery of the Guide when Bent.

From eqns. (25) and (26) we get the surface impedance Z_s at $r = r_0$ as

$$Z_s = jX_s = \frac{\mathscr{E}_\phi}{\mathscr{H}_\theta} = -\frac{j\omega\mu}{U}\frac{\mathrm{J}_1(Ur_0)}{\mathrm{J}_0(Ur_0)} \quad . \quad . \quad . \quad (28)$$

so that variation of the reactance X_s is associated with a corresponding variation of U.

If, therefore, we keep the permeability μ and the permittivity ϵ of the dielectric medium within the guide constant, we can arrange for the necessary variation of β by changing the surface reactance around the circumference of the tube when bent. The incidental change in the value of U is not large provided that $R \gg r$, but there may still be a significant disturbance of the circular H_{01} field distribution in the guide. The two conflicting requirements that both βR and U remain constant must clearly, in these circumstances, be the subject of a compromise. We have already established the importance of keeping βR constant, and it is therefore necessary, when using the technique of varying the surface reactance for this purpose, to examine the consequences of the corresponding unavoidable change in U.

When U is constant, eqn. (21) yields the desired solution $\mathscr{E}_{\phi_r} = A_2\mathrm{J}_1(Ur)$. The three terms in eqn. (21), namely

$$\frac{\partial^2\mathscr{E}_{\phi_r}}{\partial r^2} = A_2\left[\frac{2}{r^2}\mathrm{J}_1(Ur) - U^2\mathrm{J}_1(Ur) - \frac{U}{r}\mathrm{J}_0(Ur)\right] \quad . \quad (29)$$

$$\frac{1}{r}\frac{\partial\mathscr{E}_{\phi_r}}{\partial r} = A_2\left[\frac{U}{r}\mathrm{J}_0(Ur) - \frac{1}{r^2}\mathrm{J}_1(Ur)\right] \quad . \quad . \quad . \quad (30)$$

and $$\left(U^2 - \frac{1}{r^2}\right)\mathscr{E}_{\phi_r} = A_2\left[U^2\mathrm{J}_1(Ur) - \frac{1}{r^2}\mathrm{J}_1(Ur)\right] \quad . \quad (31)$$

summate to zero as required.

However, if U is a function of r and ϕ, as given by eqns. (22) and (23), so that

$$U^2 = \omega^2\mu\epsilon - \beta^2 = \omega^2\mu\epsilon - \left(\frac{\beta_a}{1 + \dfrac{r}{R_a}\cos\phi}\right)^2$$

we then find, for the three terms in eqn. (21),

$$\frac{\partial^2\mathscr{E}_{\phi_r}}{\partial r^2} = A_2\left\{\left[\frac{2}{r^2}\mathrm{J}_1(Ur) - U^2\mathrm{J}_1(Ur) - \frac{U}{r}\mathrm{J}_0(Ur)\right] + \frac{\beta^2\cos^2\phi}{UR}\right.$$
$$\left.\left[\left(\frac{6}{Ur} - 2Ur - \frac{3}{UR} - \frac{\beta^2}{U^3R}\right)\mathrm{J}_1(Ur) - \left(\frac{3r}{R} + \frac{\beta^2 r}{U^2R}\right)\mathrm{J}_0(Ur)\right]\right\} \quad . \quad . \quad . \quad . \quad (32)$$

$$\frac{1}{r}\frac{\partial\mathscr{E}_{\phi_r}}{\partial r} = A_2\left\{\left[\frac{U}{r}\mathrm{J}_0(Ur) - \frac{1}{r^2}\mathrm{J}_1(Ur)\right]\right.$$
$$\left. - \frac{\beta^2\cos\phi}{UR}\left[\frac{1}{Ur}\mathrm{J}_1(Ur) - \mathrm{J}_0(Ur)\right]\right\} \quad (33)$$

and $$\left(U^2 - \frac{1}{r^2}\right)\mathscr{E}_{\phi_r} = A_2\left[U^2\mathrm{J}_1(Ur) - \frac{1}{r^2}\mathrm{J}_1(Ur)\right] \quad . \quad (34)$$

which clearly do not summate to zero.

To maintain a near approach to the required field pattern in the guide at a bend we conclude that

(a) From a comparison of eqns. (29) and (32),

$$\left(\frac{2}{U^2r^2} - 1\right) \gg \frac{\beta^2\cos^2\phi}{U^3R}\left(\frac{6}{Ur} - 2Ur - \frac{3}{UR} - \frac{\beta^2}{U^3R}\right)$$

and $$\frac{1}{Ur} \gg \frac{\beta^2\cos^2\phi}{U^3R}\left(\frac{3r}{R} + \frac{\beta^2 r}{U^2R}\right)$$

(b) From a comparison of eqns. (30) and (33),

$$\frac{1}{U^2r^2} \gg \frac{\beta^2\cos\phi}{U^3R}\left(\frac{1}{Ur}\right)$$

and $$\frac{1}{Ur} \gg \frac{\beta^2\cos\phi}{U^3R} \quad (35)$$

The worst case is when $\phi = 0$ or 180° (i.e. on the outside or the inside of the bend), and since Ur ranges from zero up to about 4·5 in order to give the necessary surface reactance at the wall of the guide, it follows that, with $R \gg r$, we can satisfy all the above requirements in eqn. (35) when

$$\frac{\beta^2 r}{U^2R} \ll 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

For an air-filled guide with $r_0 = 1{\cdot}72$ cm (tube of $1\frac{1}{2}$ in outside

diameter) operated at a frequency of 35 000 Mc/s and bent to have a radius of curvature, R, of 5 m, say, we find, when $r = r_0$,

$$\frac{\beta^2 r}{U^2 R} = 0 \cdot 023$$

Thus, in a case of the kind postulated, where the radius of curvature is about 300 times the radius of the guide, it seems reasonable to expect that the circular H_{01}-mode would not be seriously contaminated by other modes.

(2.3.2.1) *Calculation of the Change of Surface Reactance required to produce the Appropriate Change of Phase Velocity over the Cross-Section of the Guide when Bent.*

For the circular H_{01}-mode eqn. (28) gives the surface reactance X_s for any given value of U, and eqn. (22) relates these quantities to the corresponding phase velocity, $v_p = \omega/\beta$, at the wall of the guide.

Alternatively the change of phase velocity can be calculated for a small change of surface reactance from the expression

$$\frac{dX_s}{dv_p} = -\frac{\mu\beta^3}{U^3}\left\{Ur_0 - 2\frac{J_1(Ur_0)}{J_0(Ur_0)} + Ur_0\left[\frac{J_1(Ur_0)}{J_0(Ur_0)}\right]^2\right\} \quad (37)$$

Using these relationships it is of interest to plot, as in Fig. 2, the change of phase velocity Δv_p for different values of the surface reactance of the air-filled guide, taking as an example a tube with $r_0 = 1 \cdot 72$ cm ($1\frac{1}{2}$ in outside diameter) supporting the low-loss mode at a frequency $f = 35\,000$ Mc/s. For comparison, similar curves are given for the H_{11}-mode (Fig. 2) the E_{01}-mode and the E_{11}-mode (Fig. 3) in the same guide. It is important to notice the very large surface reactances required for the H-modes compared with the E-modes, in order to bring about a given change of phase velocity.

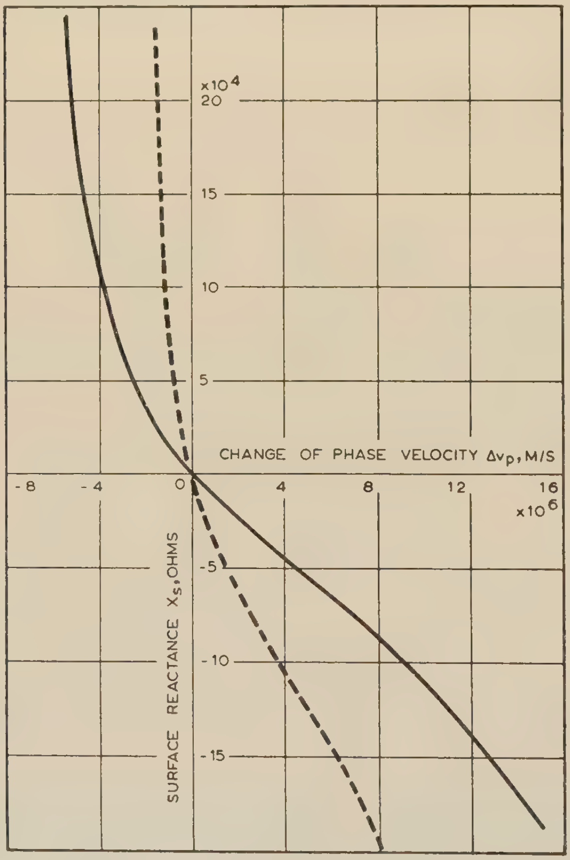


Fig. 2.—Effect of surface reactance on phase velocity for tubular air-filled guide.

$r_0 = 1 \cdot 72$ cm.
$f = 35\,000$ Mc/s.
——— H_{01} mode.
– – – – H_{11} mode.
At $X_s = 0$
$v_p = 3 \cdot 15 \times 10^8$ m/s for H_{01}-mode.
$v_p = 3 \cdot 03 \times 10^8$ m/s for H_{11}-mode.

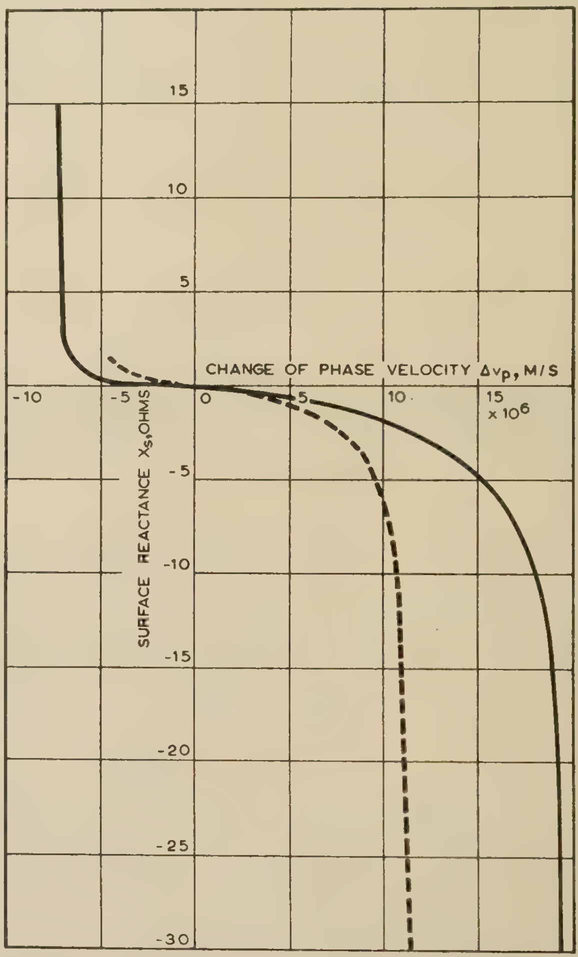


Fig. 3.—Effect of surface reactance on phase velocity for tubular air-filled guide.

$r_0 = 1 \cdot 72$ cm.
$f = 35\,000$ Mc/s.
——— E_{11}-mode.
– – – – E_{01}-mode.
At $X_s = 0$
$v_p = 3 \cdot 15 \times 10^8$ m/s for E_{11}-mode.
$v_p = 3 \cdot 06 \times 10^8$ m/s for E_{01}-mode.

To maintain a radial wavefront with respect to the centre of curvature of a bend in the guide we need a change of phase velocity between the inside and the outside of the bend given by

$$v_{p0} - v_{p1} = \left(\frac{2r_0}{R_a}\right) v_{pa}$$

where v_{pa} = phase velocity at $R = R_a$.

Thus for the circular H_{01}-mode in an air-filled guide with

$$r_0 = 1 \cdot 72 \text{ cm}$$
$$R_a = 5 \text{ m}$$
$$v_{pa} = 3 \cdot 15 \times 10^8 \text{ m/s}$$

(the value of v_p for $X_s = 0$ when $f = 35\,000$ Mc/s)

we find $(v_{p0} - v_{p1}) = 2 \cdot 17 \times 10^6$ m/s

(3) PRACTICAL CONSIDERATIONS

The preceding theoretical examination of the problem of bends in low-loss waveguide leads to the conclusion that it is desirable to bring about a change of phase velocity over the cross-section of the guide, automatically adjusted to the radius of curvature. We now have to consider a practical means of meeting this requirement, first by varying the permittivity of the dielectric medium within the tube and secondly by varying the reactance of the surface of the guide around its circumference.

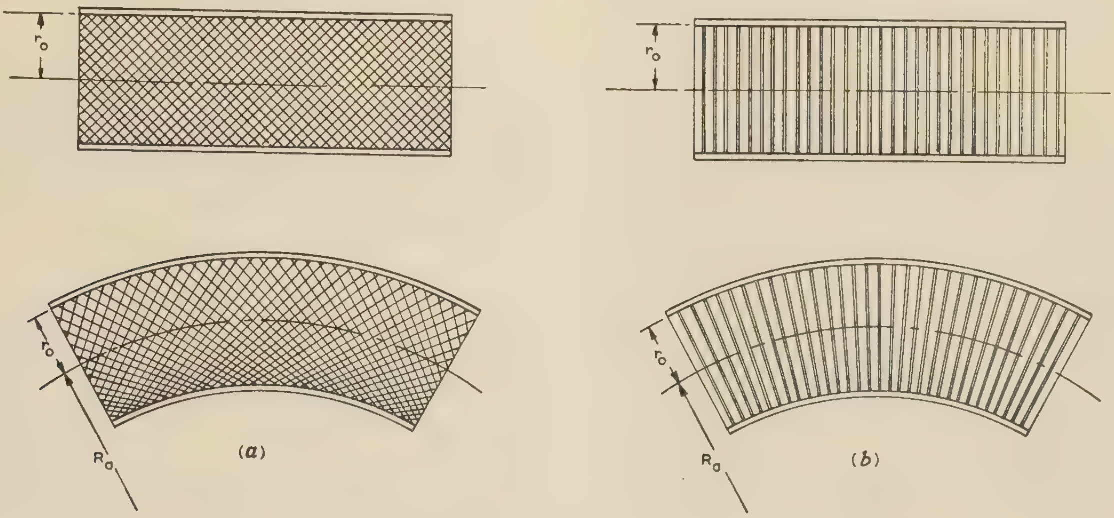


Fig. 4.—Variation of permittivity of dielectric medium inside the waveguide.

(*a*) Use of cellular polystyrene or polythene. Straight and bent sections of guide.
(*b*) Use of spaced dielectric discs. Straight and bent sections of guide.

(3.1) Variation of Permittivity of Dielectric Medium inside the Guide

Eqn. (24) gives the value of ϵ required at any point, defined by $r \cos \phi$, over the cross-section of the guide. In order to reduce losses as far as possible, it is important to fill the space inside the guide mostly with air; and cellular polystyrene or polythene capable of some mechanical compression on the inside of the bend, and further expansion on the outside, therefore suggests itself for the dielectric medium [Fig. 4(*a*)]. Alternatively, it would be possible to use a series of thin discs of polythene (cellular or otherwise) filling the cross-section of the tube and spaced by air [Fig. 4(*b*)]. In order to get an idea of the range of permittivities required for a given application, it is helpful to apply eqn. (24) to the case already discussed, in which $r_0 = 1 \cdot 72\,\text{cm}$ and $f = 35\,000\,\text{Mc/s}$. If we suppose that the relative permittivity of the dielectric in the tube at $R = R_a$ is $\epsilon_{ra} = 1 \cdot 05$ then $\beta_a = 7 \cdot 17 \times 10^2\,\text{rad/m}$ (for $X_s = 0$). When $R_a = 5\,\text{m}$ we find from eqn. (24) the relative permittivity on the outside of the bend, ϵ_{r0}, is $1 \cdot 0434$, and that on the inside of the bend, ϵ_{r1} is $1 \cdot 0566$. When $R_a = 1\,\text{m}$ we obtain $\epsilon_{r0} = 1 \cdot 017$ and $\epsilon_{r1} = 1 \cdot 083$.

It is surprising what a small change of permittivity suffices for this purpose, and there should be no great difficulty in providing for it by the method described.

(3.2) Variation of the Reactance of the Surface of the Guide around its Circumference, to the Circular H_{01} Mode

In Section 2.3.2.1 it was calculated that an air-filled guide of inside radius $1 \cdot 72\,\text{cm}$ operated at a frequency of $35\,000\,\text{Mc/s}$, and bent to a radius of curvature of 5 m would require a change of phase velocity between the inside and the outside of the bend of $2 \cdot 17 \times 10^6\,\text{m/s}$. From Fig. 2 it is seen that the easiest way to get this change of phase velocity is to use a capacitive surface for which dv_p/dX_s is larger than for the corresponding inductive surface. In any case, a comparatively large change of reactance is required, and to get this, circumferential grooves in the wall of the guide each supporting an H_{011} coaxial mode suggest themselves as most appropriate (Fig. 5). A difficulty arises in applying this device because it is necessary to make the surface reactance effectively uniformly distributed along the length of the guide,

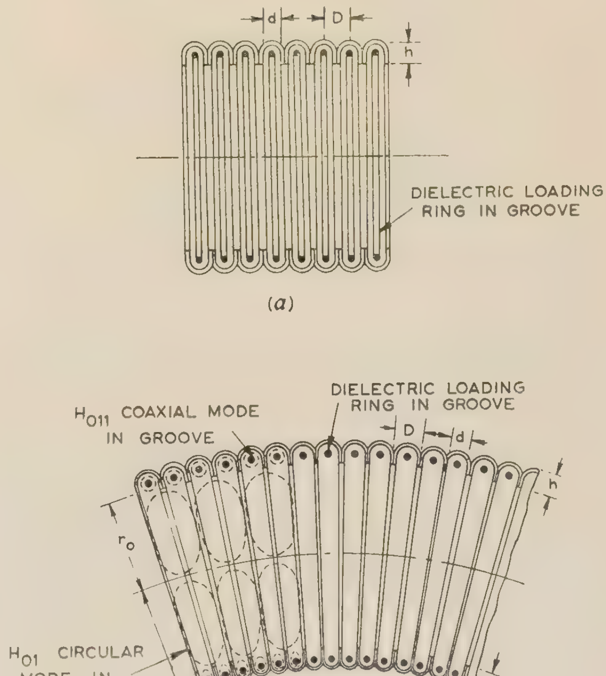


Fig. 5.—Variation of surface reactance using corrugated guide.

(*a*) Straight section.
(*b*) Bent section.

*

and that requires at least three circumferential grooves per wavelength of the circular H_{01}-mode in the main guide. Clearly this can only be achieved by loading each groove with a high-permittivity dielectric, thus enabling the groove width and pitch to be correspondingly reduced. When the guide is bent it is necessary to make the outside of the bend more capacitive and the inside less so. This can be brought about by the change in the cross-sectional dimensions of the grooves around the circumference of the tube when bent. Referring to Fig. 5 we know[6] that, if there are about three grooves or more within the wavelength of the main guide,

$$X_s = X_g \frac{d}{D} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

where X_g is the groove reactance.

For a coaxial groove partially filled with solid dielectric and supporting the H_{011}-mode, a precise calculation for X_g is difficult to make, but an approximate estimate shows that a change of reactance of the order required is readily obtained with the change in d, and particularly in h, that takes place when the guide is bent. The solid dielectric in the groove should preferably have a relative permittivity of about 8 with small losses (e.g. polythene loaded with titanium dioxide) and be placed as a coaxial ring at the mean radius. Since currents in the wall of the groove are entirely circumferential, a guide, fabricated from a series of metal rings mounted inside a rubber tube and separated from one another by thin flexible metal spacers each slightly in compression whilst forming the bottom of a groove, can be assembled as shown in Fig. 6. Alternatively a thin-walled metal tube on which circular corrugations have been formed, either by spinning or by applying an internal fluid pressure against an external die, can be used, but in this case the loading of the grooves presents more difficulty.

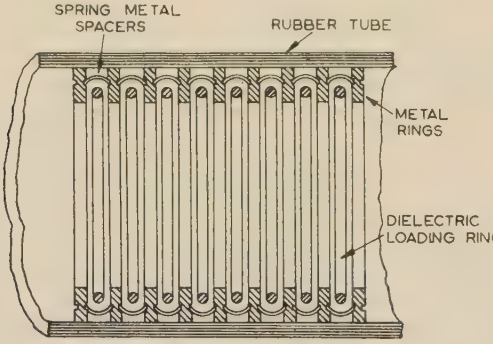


Fig. 6.—Fabricated corrugated guide for variation of surface reactance.

It is perhaps relevant to recall the recent announcement detailed in Reference 5 of the application of a spiral wire guide to deal with bends. As indicated in Fig. 7, it seems that this form of construction may be arranged to provide for the appropriate kind of field distribution at the inside surface, but to get the range of reactance at that surface and to control it as suggested here would appear to be a very difficult matter. Moreover, the spiral wire formation is bound to introduce some longitudinal electric field, however small, and this is undesirable.

(3.2.1) Other Uses for Circular Corrugations.

As already shown in Reference 4, a circularly corrugated guide can be usefully employed in the low-loss application (*a*) to act as a filter for unwanted modes, (*b*) to absorb energy in unwanted modes and (*c*) to break the degeneracy between the circular H_{01}-mode and the E_{11}-mode which tends to appear at bends.

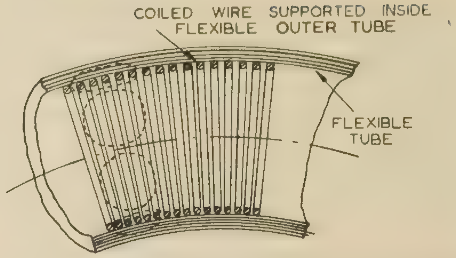


Fig. 7.—Flexible waveguide for circular H_{01}-mode proposed in Reference 5.

These are all important techniques and will no doubt find a place in any practical development, particularly because they do not involve putting anything into the main channel of the guide. Nevertheless, the fact that these devices have been considered necessary is an indication of the lack of complete success previously achieved in dealing with this problem.

(4) CONCLUSIONS

Analysis of the problem of bends in low-loss waveguide leads to the important conclusion that the phase velocity at any point over the cross-section of the tube must be so adjusted that the wavefront always remains radial with respect to the centre of curvature. It is suggested that there are two practical ways of approaching this requirement: (*a*) by suitable variation of the permittivity of the dielectric medium inside the guide over its cross-section, and (*b*) by varying the surface reactance of the guide around its circumference, in relation to the circular H_{01}-mode. Proposal (*a*) seems to be not only simpler and more easily applied but also likely to be more effective because it provides for control of the phase velocity over the whole wavefront. Experiments are now in hand to test the validity of the conclusions reached.

(5) ACKNOWLEDGMENTS

The author wishes to thank his colleagues, Dr. John Brown and Mr. H. G. Effemey, for helpful discussion of this problem, and Mr. O. E. Fletcher of Imperial Chemical Industries, Ltd., and Mr. R. C. Mildner of the Telegraph Construction and Maintenance Co., Ltd., for their encouragement in pursuing it. He is also indebted to the Admiralty Signals Establishment, whose support of experimental work on the subject, now being carried out at University College, London, by Messrs. M. G. F. Wilson and D. C. Rickard, has helped to stimulate further interest.

(6) REFERENCES

(1) Barlow, H. E. M., and Effemey, H. G.: 'Propagation Characteristics of Low-Loss Tubular Waveguides', *Proceedings I.E.E.*, Paper No. 2326 R, May, 1957 (**104 B**, p. 254).

(2) Jouguet, M.: 'Wave Propagation in nearly Circular Waveguides: Transmission-over-Bend Devices for H_0-Waves', *Câbles et Transmission*, 1948, **2**, p. 257.

(3) Sims, G. D.: 'The Influence of Bends and Ellipticity on the Attenuation and Propagation Characteristics of the H_{01} Circular Waveguide Mode', *Proceedings I.E.E.*, Monograph No. 58 R, January, 1953 (**100**, Part IV, p. 25).

(4) Miller, S. E.: 'Notes on Methods of Transmitting the Circular Electric Wave around Bends', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1104.

Miller, S. E., and Beck, A. C.: 'Low-Loss Waveguide Transmission', *ibid.*, 1953, **41**, p. 348.

Albersheim, W. J.: 'Propagation of TE_{01} Waves in Curved Waveguides', *Bell System Technical Journal*, 1949, **28**, p. 1.

(5) Announcement by Bell Telephone Laboratories: 'A New Approach to Long-Distance Transmission', *Radio Electronics*, April, 1955, p. 45; and *Engineers Digest*, May, 1955, p. 184. See also MORGAN, S. P., and YOUNG, J. A.: 'Helix Waveguide', *Bell System Technical Journal*, 1956, **35**, p. 1347. (The paper deals with straight guides only.)
(6) BARLOW, H. E. M., and KARBOWIAK, A. E.: 'An Experimental Investigation of the Properties of Corrugated Cylindrical Surface Waveguides', *Proceedings I.E.E.*, Paper No. 1625 R, May, 1954 (**101**, Part III, p. 182).

(7) APPENDIX

In order to get a better idea of the conditions imposed and the error introduced by assuming that the field components $\mathscr{E}_\phi$, $\mathscr{H}_r$ and $\mathscr{H}_\theta$ suffice to satisfy Maxwell's equations for the curved guide, the following analysis is helpful:

If $\mathscr{E}_\phi = \mathscr{E}_{\phi a}$, $\mathscr{H}_\theta = \mathscr{H}_{\theta_a}$ and $\beta = \beta_a$ when $R = R_a$ (i.e. $\phi = 90°$ or $270°$), eqns. (8), (9) and (11) give, without approximation,

$$\mathscr{E}_{\phi_a} = A\varepsilon^{-j\beta_a R_a \theta} J_1(Ur) \quad . \quad . \quad . \quad . \quad (39)$$

and

$$\mathscr{H}_{\theta_a} = -\frac{UA}{j\omega\mu}\varepsilon^{-j\beta_a R_a \theta} J_0(Ur) \quad . \quad . \quad . \quad (40)$$

We also have from eqn. (12) after integration

$$\mathscr{H}_\theta = \frac{R_a}{R}\mathscr{H}_{\theta_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Using eqn. (40), we therefore find

$$\mathscr{H}_\theta = -\frac{UAR_a}{j\omega\mu R}\varepsilon^{-j\beta_a R_a \theta} J_0(Ur) \quad . \quad . \quad . \quad (42)$$

and from eqn. (8) we get

$$\mathscr{E}_\phi = \frac{AR_a}{r}\varepsilon^{-j\beta_a R_a \theta}\int \frac{Ur}{R} J_0(Ur)dr \quad . \quad . \quad . \quad (43)$$

The constant of integration in eqn. (43) is zero because this expression reduces to eqn. (39) when $R = R_a$.

Comparing eqns. (42) and (43) with eqns. (26) and (25), respectively, it is clear that the condition

$$\beta R = \beta_a R_a = \text{constant}$$

is of paramount importance. Furthermore, eqn. (42) differs very little from eqn. (26) when $R \gg r$. Eqn. (43) can be rewritten as

$$\mathscr{E}_\phi = \frac{A}{r}\varepsilon^{-j\beta_a R_a \theta}\int\left(1 - \frac{r\cos\phi}{R_a} + \frac{r^2\cos^2\phi}{R_a^2}\dots\right)UrJ_0(Ur)dr \quad (44)$$

and the first term in this expression after integration is the same as (25).

From eqns. (9) and (43), we find

$$\mathscr{H}_r = -\frac{AR_a^2\beta_a}{\omega\mu Rr}\varepsilon^{-j\beta_a R_a \theta}\int \frac{1}{R}UrJ_0(Ur)dr \quad . \quad . \quad (45)$$

and clearly, therefore, eqn. (10) is not precisely satisfied, although when $R \gg r$ the error is small.

621.385.3: 621.365.5

Paper No. 2379 R
July 1957

THE DESIGN AND OPERATION OF HIGH-POWER TRIODES FOR RADIO-FREQUENCY HEATING

By W. J. POHL, M.Sc., Associate Member.

(*The paper was first received* 1*st October and in revised form* 3*rd December*, 1956, *and in final form* 25*th March*, 1957.)

SUMMARY

The paper sets out the important factors in the design of high-power triodes for use in radio-frequency heating equipment. The relations between Class C oscillator performance and valve characteristics lead to relatively simple design principles. These, in combination with modern methods of construction, have made it possible to achieve an unusually high conversion efficiency with relatively low dissipation on the grid, thus ensuring adequate grid safety under varying load conditions.

The design principles are embodied in a new range of industrial triodes which are briefly described.

LIST OF SYMBOLS

d = Diameter of grid wires.
V = Direct supply voltage.
I_k = Peak instantaneous total current per unit area.
I_a = Peak instantaneous anode current per unit area.
I_g = Peak instantaneous grid current per unit area.
I_{dg} = Direct grid current per unit area.
I_{1a} = Peak value of fundamental component of anode current per unit area.
I_{1g} = Peak value of fundamental component of grid current per unit area.
l_a = Grid-to-anode distance measured to the centre of the wires.
l_g = Grid-to-cathode distance measured to the centre of the wires.
p = Pitch of grid wires.
V_g = Peak value of instantaneous grid voltage relative to cathode.
V_a = Minimum instantaneous anode voltage relative to cathode.
$P_{g\,max}$ = Maximum permissible grid dissipation per unit area of grid.
P_g = Total grid dissipation per unit area of grid.
μ = Triode amplification factor.
ϕ = Angle of grid current flow, electrical degrees.
V_s = Peak value of grid voltage relative to bias voltage.
V_b = Bias voltage.
P_d = Drive power (fed back from the ouput in an oscillator).

(1) INTRODUCTION AND SYNOPSIS

In the increasingly important field of radio-frequency heating, high-power triodes are generally used as Class C oscillators. The circumstances which prevail in most cases demand that the valve should give satisfactory service over a very wide range of r.f. load conditions.

Unfortunately, a Class C oscillator which has been adjusted for high efficiency of operation at full load will, under no-load conditions, run with excessive grid drive, frequently followed by grid failure. The power dissipated on the grid at no-load may be several times that which is dissipated at full load. It is therefore necessary that the full-load conditions be arranged so that the grid dissipation is low enough to allow an adequate safety margin.

In this paper it is shown that the grid dissipation is very critically dependent on the operating conditions. As the full-load efficiency of the oscillator is increased, the grid dissipation rises very sharply. Thus, the circuit engineer who, by adjustment of feedback and bias, derives the utmost power from his circuit, jeopardizes the valve grid safety. It is proposed here to use the concept of a 'grid safety factor', defined as the ratio of maximum rated grid dissipation to actual grid dissipation at full load.

It would perhaps have been more logical to define this factor with reference to no-load conditions. Since, however, these depend not only on the valve but also on the circuit losses, and in any case are difficult to calculate, it is assumed for the purposes of the paper that the safety factors at full load and at no-load bear a constant relationship.

The valve designer is faced with the task of providing an acceptable grid safety factor at the highest possible anode efficiency. By examining the interdependence of these two parameters for various designs, it will be shown that this requirement is best met by the smallest mechanically reliable grid-cathode spacing (high slope), and the smallest grid-anode spacing (giving low amplification factor) compatible with freedom from flashover. The effects of variations in wire diameter and grid pitch are also considered, and it is shown that, although it is advantageous to use fine wire, it is not worth while to make great efforts to reduce the wire diameter below that which is mechanically convenient.

These principles, together with modern methods of construction, have been applied in the development of a new range of industrial-heating triodes, and have resulted in considerable improvements in efficiency and grid safety factor.

(2) CLASS C OSCILLATOR RELATIONS

The performance equations describing Class C oscillator conditions will now be derived. These relations enable the circuit designer to carry out Class C calculations far more rapidly than by conventional methods.

A flow angle of 140° will be used, this angle being considered representative of full-load operation, with good compromise between adequate output power and efficiency. The flow angle merely determines the constants in the equations below, which are obtained by Terman's method.[1] Shorter flow angles lead to higher efficiencies at lower powers, but it will be seen that the generalized conclusions concerning valve design which are derived later are equally applicable to any angle of flow.

The available radio-frequency power is

$$\left(\frac{V - V_a}{2}\right) I_{1a} = 0{\cdot}2 I_a(V - V_a) \quad . \quad . \quad . \quad (1)$$

The drive power is

$$P_d = \frac{I_{1g} V_s}{2} = \frac{I_{1g}}{2}(V_b + V_g)$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Pohl is with Standard Telephones and Cables, Ltd.

where $$V_b = \frac{V}{\mu} + 0{\cdot}52\left(V_g + \frac{V_a}{\mu}\right) \quad . \quad . \quad . \quad . \quad (2)$$

Hence the drive power is

$$\frac{I_{1g}}{2}\left[\frac{V}{\mu} + 0{\cdot}52\left(V_g + \frac{V_a}{\mu}\right) + V_g\right]$$

The ratio V_a/μ is generally small, so that

$$P_d = \frac{I_{1g}}{2}\left(\frac{V}{\mu} + 1{\cdot}6V_g\right) \quad . \quad . \quad . \quad . \quad (3)$$

is a good approximation, where $V_s = V/\mu + 1{\cdot}6V_g$.

It is shown in Section 10.1 that for practical purposes the grid flow angle, in radians, can be taken as

$$\phi = 2\sqrt{2}\sqrt{\frac{V_g}{V_s}} = 2\sqrt{2}\sqrt{\frac{V_g}{V/\mu + 1{\cdot}6V_g}} \quad . \quad . \quad (4)$$

and that $I_{1g}/I_g = 0{\cdot}17\phi$ is an adequate approximation. Hence the drive power is

$$P_d = \frac{0{\cdot}17I_g}{2}2\sqrt{2}\sqrt{\left(\frac{V_g}{V_s}\right)}V_s = 0{\cdot}24I_g\sqrt{[V_g(V/\mu + 1{\cdot}6V_g)]} \quad (5)$$

Subtracting this from the available radio-frequency power one obtains the output power:

$$P_o = 0{\cdot}2I_a(V - V_a) - 0{\cdot}24I_g\sqrt{[V_g(V/\mu + 1{\cdot}6V_g)]} \quad (6)$$

An expression for the grid dissipation is derived in Reference 2 and is

$$P_g = \frac{\phi}{4\pi}V_gI_g \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Eqns. (4) and (7) then yield a useful expression for the grid dissipation in terms of peak grid voltages and currents as obtained from measured or calculated valve characteristics,

$$P_g = 0{\cdot}225\sqrt{\left(\frac{V_g}{V_s}\right)}V_gI_g = 0{\cdot}255\sqrt{\frac{V_g}{(V/\mu + 1{\cdot}6V_g)}}V_gI_g \quad (8)$$

which is a satisfactory approximation for grid flow angles up to about 120°. (See also Section 10.2.) The anode efficiency is given by the quotient of output power and direct input power, i.e.

$$\eta = \frac{P_o}{P_{di}} = \frac{4{\cdot}4P_o}{I_aV} \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

To find the bias resistance it is necessary to know the direct grid current I_{dg}, which from Reference 1 is approximated by

$$I_{dg} = 0{\cdot}09\phi I_g$$

Substituting for ϕ from eqn. (4),

$$R_{bias} = \frac{V_s - V_g}{0{\cdot}254I_g}\sqrt{\frac{V_s}{V_g}} \quad . \quad . \quad . \quad . \quad (10)$$

Summary.—The design is carried out by first considering eqn. (1). For any assumed value of V_a the characteristic curves of the valve give V_g and I_g for the required current I_a.

The drive power required and the grid voltage swing, V_s, are derived from eqn. (3). Eqns. (6), (8), (9) and (10) then give output, grid dissipation, efficiency and bias resistance, respectively. It may, of course, be expedient to provide some of the bias voltage by means of a cathode bias resistor.

It should be noted that the above procedure applies to earthed-cathode operation. At high frequencies (in the region of 100 Mc/s) it will become necessary to employ the earthed-grid method.[2] For this it is necessary only to replace I_{1g} in eqn. (3) by $I_{1g} + I_{1a}$. The grid dissipation remains unchanged.

The constants in the equations can be readily modified with the aid of Reference 1 if an anode flow angle other than 140° is required. In general, it will be found that smaller flow angles give higher efficiencies at lower output powers for given filament emission capabilities. This means that if a valve is under-run very high efficiencies can be obtained. An example of this is given in the measured test results in Table 1 where the valve type 3J222E, which is capable of 24 kW output power, gives an efficiency in excess of 90% when only 17 kW of power are required.

Table 1

Typical Measured Oscillator Performance Data

	Low-voltage types	High-voltage types
	3J/222E 3Q/222E	3J/252E 3Q/252E
Output power (measured into load)	24 kW	48 kW
Grid dissipation	120 W	200 W
Efficiency	78%	80%
At direct anode voltage ..	6·0 kV	12·0 kV
direct anode current ..	5·1 A	5·0 A
direct grid current	0·71 A	0·8 A
grid bias resistance	800 Ω	950 Ω
peak positive grid voltage ..	200 V	320 V
	Low-voltage types	High-voltage types
	3J/202E 3Q/202E	3J/232E 3Q/232E
Output power (measured into load)	12 kW	24 kW
Grid dissipation	60 W	100 W
Efficiency	78%	80%
At direct anode voltage ..	6·0 kV	12·0 kV
direct anode current ..	2·6 A	2·5 A
direct grid current	0·36 A	0·4 A
grid bias resistance ..	1 600 Ω	1 900 Ω
peak positive grid voltage ..	200 V	320 V

Owing to their special characteristics, these valves show an unusually sharp increase in efficiency if the output power requirements are slightly reduced. For example, the 3J/222E at 17 kW output power gives 90% efficiency, with only 50 watts grid dissipation.

Output power (measured into load) ..	17 kW	Grid current	0·32 A
Bias resistance ..	2 500 Ω	Peak positive grid voltage	200 V
Anode current ..	2·9 A	Grid dissipation ..	50 W
Anode voltage ..	6·5 kV	Efficiency	90%

Measured at approximately 2 Mc/s with less than 5% circuit loss.

(3) CALCULATION OF VALVE CHARACTERISTICS

To establish design principles, expressions are required relating valve-electrode dimensions to the peak currents and voltages used in the previous Section.

A widely used relationship for the total space current is

$$I_k = \frac{2{\cdot}34 \times 10^{-6}(V_g + V_a/\mu)^{3/2}}{\left(l_g + \frac{l_g + l_a}{\mu}\right)^2} \quad . \quad . \quad (11)$$

per unit area of cathode.

For the grid current, the approximate relationship for a parallel-wire grid

$$I_g = I_a\left(\frac{l_g + l_a}{l_g}\right)^{2/3}\frac{d}{p-d}\sqrt{\frac{V_g}{V_a}} \quad . \quad . \quad . \quad (12)$$

is used. This is derived by assuming the experimentally verified relation

$$I_g/I_a \propto \sqrt{(V_g/V_a)}$$

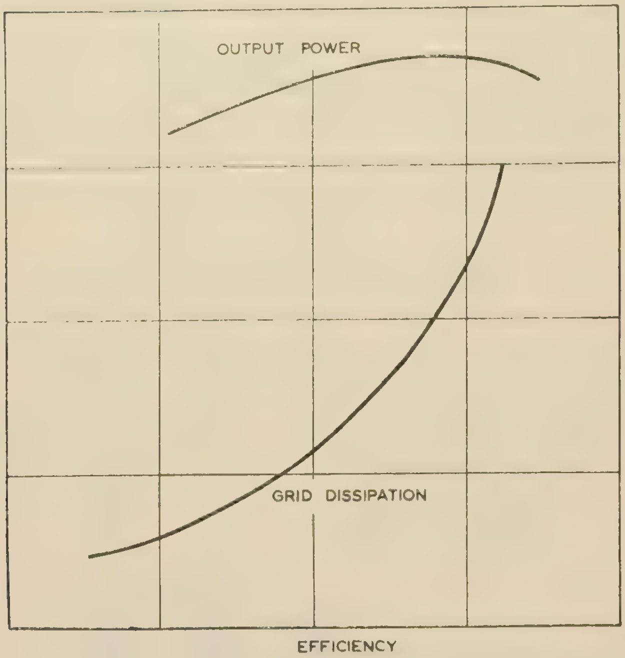


Fig. 1.—General form of performance curves.

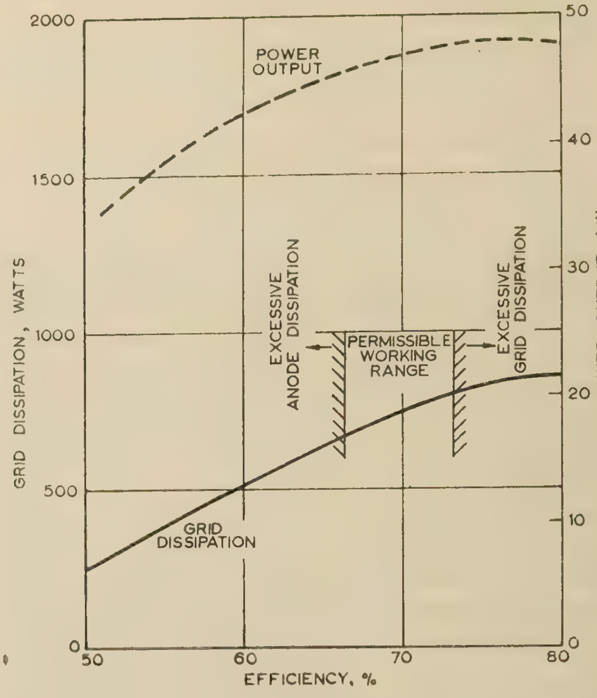


Fig. 2.—Performance curves of an obsolete triode unsuited for radio-frequency heating.

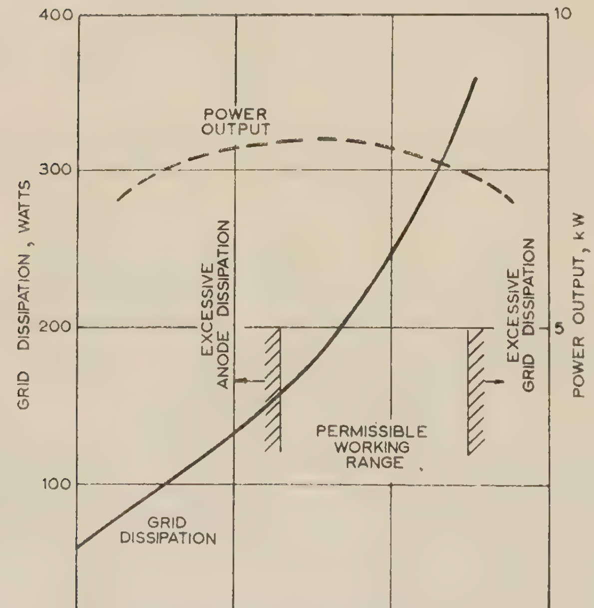


Fig. 3.—Performance curves of a triode satisfactory for radio-frequency heating.

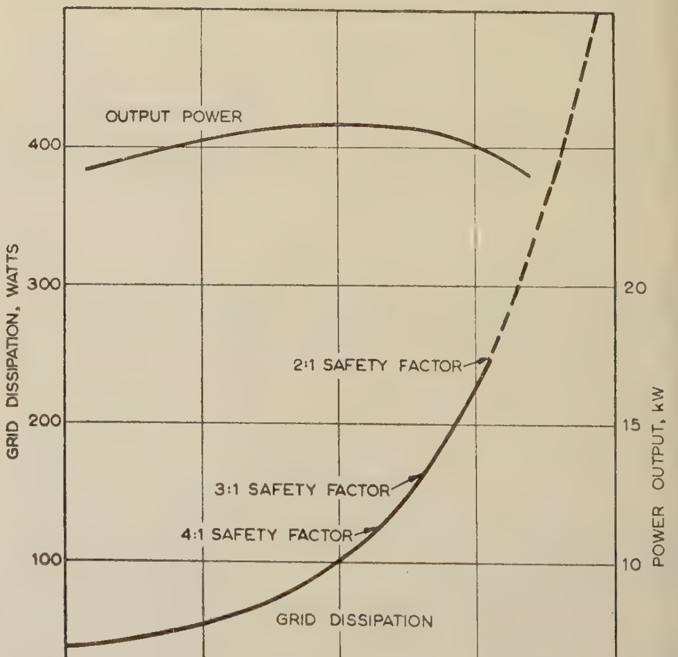


Fig. 4.—Performance curves of the 3J/222E, a triode specially designed for radio-frequency heating.

and finding the constant of proportionality.[2] Eqn. (12) is not highly accurate and has on occasions given rise to errors up to 20%. Nevertheless it suffices for use here, since we are looking for general design trends and seeking to achieve reductions in grid dissipation by an order of magnitude. (The characteristics of the square-mesh grid correspond to those of a parallel-wire grid having a pitch equal to 0·6 times that of the mesh.[6])

The calculation of μ has been carried out from information given by Ollendorf and reprinted in Reference 3.

(4) A METHOD OF ESTIMATING VALVE SUITABILITY

The foregoing relations will now be used to bring out something about the factors which are important in the design of high-power triodes for oscillators in industrial heating equipment. To this end a method of estimating valve suitability is employed which has proved useful, not only for exposing the shortcomings of some existing obsolescent types and accounting for failures in the field, but also in establishing the valve design principles outlined in the following Section. The method consists of calculating the output power, efficiency and grid dissipation at a stipulated anode voltage, flow angle (e.g. 140°) and a *fixed peak cathode current.* This gives curves of which the generalized form is shown in Fig. 1. Such curves are obtained from measured or calculated valve characteristics as follows:

A number of values of V_a are assumed. For each of these the peak grid voltages V_g required to give the fixed specified peak total current are found from the characteristic curves (preferably constant-current curves), together with the corresponding values

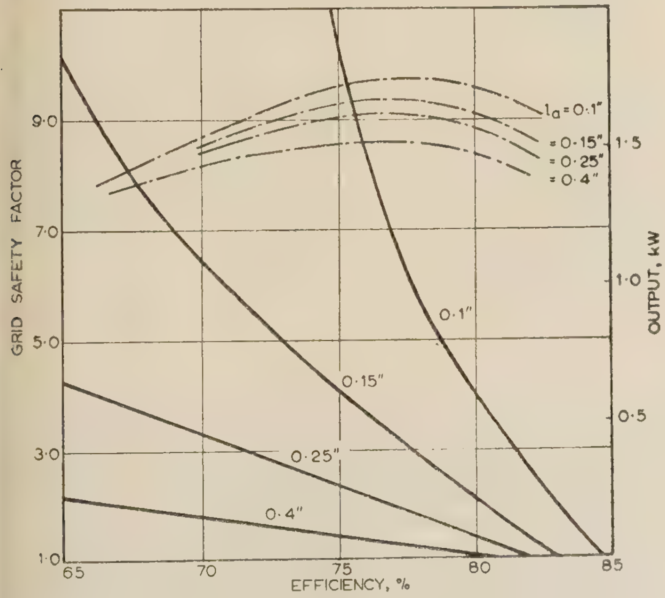


Fig. 5.—Effect of varying grid-anode spacing on grid safety factor and output power.

Safety factor ——. Output power – – –.

These curves have been calculated for one square inch of electrode area, with $V = 6$ kV, $l_g = p = 0 \cdot 060$, $I_a + I_g = 1 \cdot 6$ amp. Flow angle = 140°.

$P_{g\,max}$ is taken at 40 watts/in² of wire area, i.e. 12·8 watts for these curves.

of I_g and I_a. Eqns. (6) and (8) are then used to obtain output power and grid dissipation for each assumed value of V_a; these are readily expressed in terms of anode efficiency from eqn. (9). At efficiencies above about 60% the output power will be found to remain substantially constant, and represents the maximum power which the valve will give at the voltage considered and at the available total cathode emission. The grid dissipation will be seen to increase rapidly with efficiency. The maximum attainable efficiency is limited by the point at which the safe grid dissipation is exceeded.

As mentioned in Section 1, it is desirable, for variable-load operation, to work with a certain grid safety factor at full load. Thus a 'merit figure' of a valve could, for this aspect, be specified as the 'efficiency at full output power at which the grid has a given safety factor'. If this factor were, say, fixed at 2·0, then the valve relating to Fig. 2 gives a 'merit figure' of 56%. The

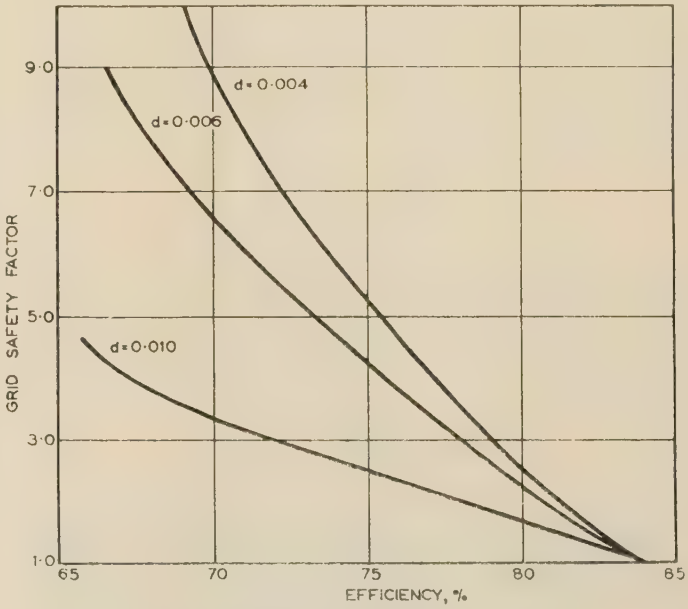


Fig. 6.—Effect of varying wire diameter.

These curves have been calculated as those of Fig. 5, and with $l_a = 0 \cdot 15$ in.

information given here applies to an obsolescent broadcast valve which is unsuitable for industrial heating, and at 56% efficiency the anode dissipation would be exceeded. A safety factor of 2·0 is therefore not possible with this valve at full output power. In the same way Fig. 3 gives a 'merit figure' of 65% and applies to a valve which has been satisfactory in the field. The valve to which Fig. 4 relates gives a figure of 81%, and shows what has been achieved with the use of design principles outlined in the following Section and modern methods of construction. It represents the 3J/222E 20 kW triode. A description of this and similar valves is given in Section 7.

Figs. 5 and 6 show some results obtained by applying the method to a hypothetical valve with calculated characteristics. These lead to important design conclusions which are discussed in detail in the following Section.

(5) CONCLUSIONS CONCERNING ELECTRODE DESIGN

Diagrams such as Figs. 1–6 reveal immediately the causes of unsatisfactory service in industrial heating applications and indicate desirable trends in design, which will now be discussed with reference to each of the important electrode dimensions.

(5.1) Filament

It is well known that the emission life of the cathode is dependent on its running temperature. In the case of thoriated-tungsten filaments the relations between running temperature, emission and life are well known.[5] In industrial heating applications, flow angles much less than 140° are sometimes used to give improved efficiencies. In general, it is recommended that a

flow angle of 90° should be catered for, giving a peak current of approximately six times the mean space current. Allowances must also be made for deterioration of emission during life and for filament-voltage fluctuations.

(5.2) Grid-Cathode Spacing

The choice of the smallest practicable grid-cathode spacing can be justified by the need to minimize the grid power, which is proportional to $V_g I_g$. For a given space current and current division, I_g remains constant. The value of V_g required to produce this space current diminishes with l_g, in accordance with eqn. (11). The grid-cathode spacing should therefore be as small as is practicable from a mechanical viewpoint; it is determined by the accuracy obtainable in the cathode and grid structures, their supports, and the overall dimensions.

(5.3) Grid Pitch

Too many grid wires give an unnecessarily large intercepted current and thus decrease the anode current, and hence the power. The grid safety factor is not affected because additional wires can dissipate additional power. If the grid pitch is too large, however, this gives rise to 'island' effects and 'high tail' characteristics leading to low efficiencies. A grid pitch equal to $0\cdot7 l_g$ is a good compromise.[4]

(5.4) Grid-Anode Spacing

The effect of varying the grid-anode spacing, keeping all other parameters constant, is shown in Fig. 5. It is clear that in order to obtain good grid safety factors at the highest efficiencies, the ratio of grid dissipation to efficiency should be as small as possible. The conclusions drawn from Fig. 5 indicate that it is desirable to reduce the grid-anode spacing until the tendency for flashover to take place enters into consideration. Doubts arise because the resulting low μ is usually associated with a high drive power. The drive power is, however, not as high as might at first be expected because it depends not only on μ but also on V_g and I_g, in accordance with eqn. (5). For a given space current I_g and V_g are considerably reduced by reducing l_a. Thus the drive power increases only slightly as l_a is diminished below values which are normally in use. In any case, the increase in drive power due to low μ is diverted into the bias resistor, and not on to the grid. As regards efficiency, the drive power is unimportant since, even with very low values of μ, it represents a very small fraction of the available radio-frequency power, in grid-driven circuits.

(5.5) Effect of Wire Diameter

Since the grid dissipation is proportional to $V_g I_g$ and the maximum rated dissipation is proportional to d/p, the safety factor is proportional to $d/(pI_g)$.

From eqn. (12) I_g is proportional to $d/(p - d)$. Hence the grid safety factor is proportional to $(p - d)/p$. In Fig. 6 the small effect of wire size on safety factor caused by change of μ, which has been neglected in the above arguments, is also taken into account. The curves show the effect of using different wire sizes in a design with $l_g = 0\cdot060$ in and $p = 0\cdot060$ in. They demonstrate the general principle that the wire diameter should be as small as possible, but that little is gained by reducing it when it is already very much smaller than the pitch.

(7) CHARACTERISTICS AND PERFORMANCE OF A NEW RANGE OF INDUSTRIAL TRIODES

The design principles outlined in the previous Section have been applied in the development of a new range of power triodes, two of which are shown in Fig. 7. The 3J/222E is designed for a useful output of 20 kW at an anode voltage of 6 kV. Its characteristics are shown in Fig. 8, and its calculated performance and grid safety factor for flow angles of 140° are given in Fig. 4. These are in good agreement with measured performance figures, some of which are given in Table 1.

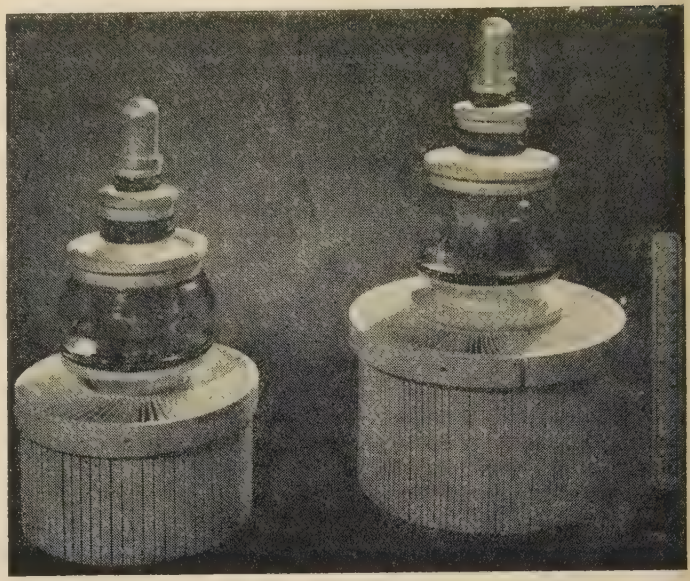


Fig. 7.—The 3J/202E and 3J/222E (10 kW and 20 kW output) industrial triodes.

The 3J/202E gives roughly half the current and power at the same voltages, and so has the same efficiencies and safety factors on the grid at full load. The 3J/232E and the 3J/252E types are designed for operation at higher voltages. The anode-grid spacing in these valves is increased in proportion to the direct supply voltage, whereas the cathode-grid spacing is maintained at the same value. This gives about the same relationship between grid safety factor and efficiency. This can be appreciated by inspection of eqns. (11) and (8) and remembering that μ is proportional to the grid-anode spacing.

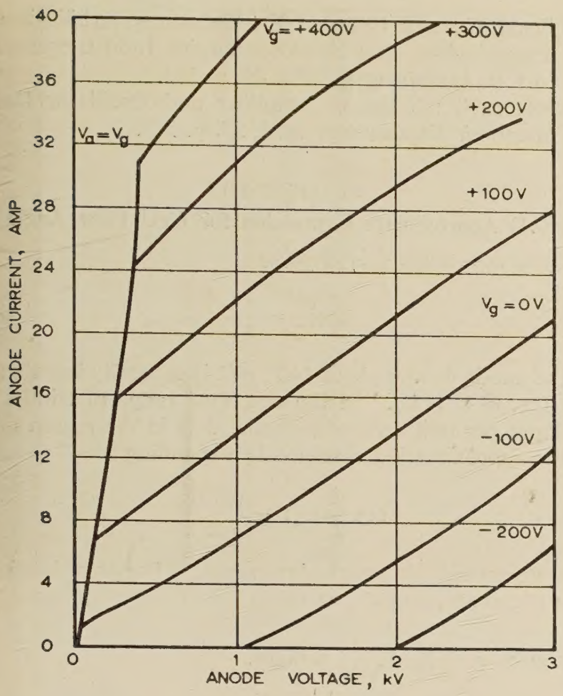
(7) MECHANICAL FEATURES OF THE NEW TRIODES

(7.1) The Envelope

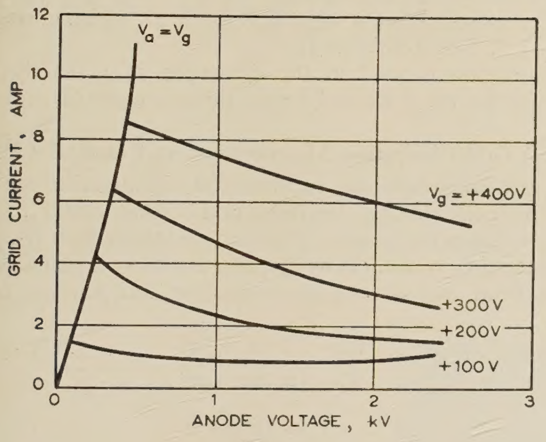
One of the most important requirements for valves used in productive industry is mechanical robustness both of the envelope and of the electrode structures. Although conventional pin-type valve connectors are adequate for the frequencies used in the majority of industrial heating applications, recent developments in the broadcasting field have pointed to the necessity of coaxial structures for higher frequencies. In this connection it was found during extensive trials that such structures not only can give reduced electrical losses and facilitate cooling, but also are many times stronger mechanically, and exceptionally resistant to thermal shock. Fig. 9 shows details of the construction used.[10] This construction facilitates accurate electrode alignment and, being entirely symmetrical, is free from warping during life. There are no vacuum-tight brazed joints, and the glass seals are of a type acknowledged to allow the widest latitude in residual stresses. Thus the envelope can be manufactured with a high degree of reliability.

(7.2) Grid

The square-mesh grid structure is shown in Fig. 10. The wires are specially processed so that at the maximum rated dissipation the emission from the grid is of the order of $2\,\mu A/cm^2$ of grid-



(a)



(b)

Fig. 8.—Characteristics of the 3J/222E.

(a) Anode-current/anode-voltage curve.
(b) Grid-current/anode-voltage curve.

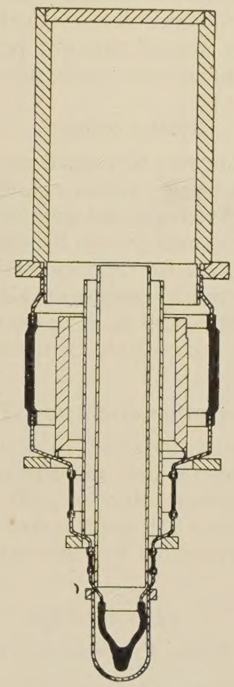


Fig. 9.—Envelope structure.

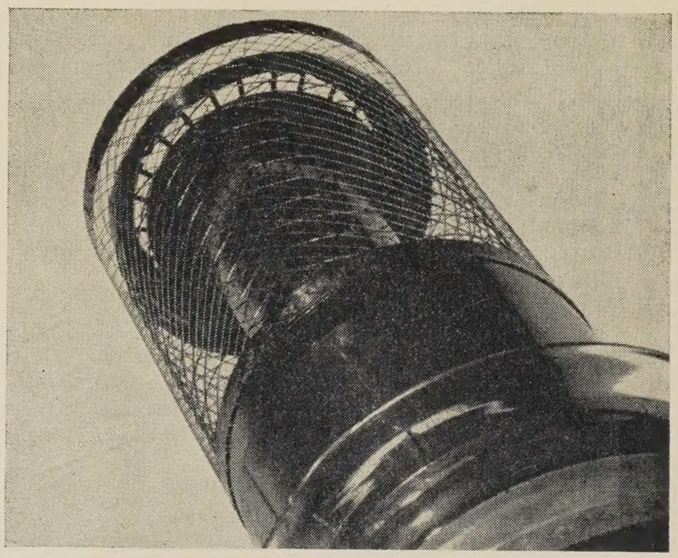


Fig. 10.—Grid-cathode assembly.

wire surface area. This occurs at about 8 watts/cm^2. The maximum rated value of grid dissipation is the figure which will be certain to ensure freedom from grid emission troubles. The grid itself, however, is capable of withstanding for short periods about four times more power before damage results.

The characteristics of the square-mesh grid correspond to those of a parallel-wire grid having a pitch equal to 0·6 times that of the mesh.[6]

(7.3) Filament

The filaments of these valves have a mesh construction. Structures such as this were first described and successfully put into practice by Müller.[7] The construction used here differs from these in so far as it is rigidly supported from one end only; this was found to give constructional advantages. The mesh consists of thoriated-tungsten wire made from two layers of multifilar helices. An important merit of this structure is that the resistance to distortion is exceptionally great, which not only makes for mechanical robustness but also allows a smaller grid-cathode spacing to be used. Electrically, the filament resembles a unipotential emitting surface rather than a collection of single filament strands with the same heating power. This characteristic and the mutual heating effect[8] between adjacent wires give an economy in filament power for a given peak current requirement. The filament is supported at one end by an inverted sleeve, and at the other is connected to a series of tapes which are shown in Fig. 10. These are designed to give the maximum freedom to the filament in the transverse and axial directions, thus minimizing mechanical strain. A central tube is used for support and for conducting the heating current to the tapes.

This structure[11] has shown a hitherto unattainable resistance to distortion. Even if the filament is burnt out by applying approximately four times the normal filament power, no movement comparable to the grid-filament spacing takes place.

(7.4) Cooling

The anode dissipation can be carried away by water, forced air or evaporative cooling.[9] Valves equipped for evaporative cooling have special circumferential grooves cut into the anode surface,[12] permitting for short periods the dissipation of twice the power considered adequate for most industrial purposes.

The radiators for forced-air cooling are also capable of double-power overloads for short periods, since the fin temperatures can under these conditions exceed 300° C without permanent damage to the valve.

(8) ACKNOWLEDGMENTS

The author is greatly indebted to a number of helpers, especially Mr. G. A. Morton, who made prominent contributions to the mechanical development work, Mr. D. C. Rogers, for helpful criticism, and Mr. C. H. Foulkes and Standard Telephones and Cables Ltd. for permission to publish the paper.

(9) REFERENCES

(1) TERMAN, F. E.: 'Radio Engineering' (McGraw-Hill, 1932), p. 325 *et seq.*
(2) POHL, W. J., and ROGERS, D. C.: 'U.H.F. Triodes. Design for Class C Earthed-Grid Operation', *Wireless Engineer*, 1955, **32**, p. 47.
(3) FREMLIN, J. H., HALL, R. N., and SHATFORD, P. A.: 'Triode Amplification Factors', *Electrical Communication*, 1946, **23**, p. 426.
(4) FREMLIN, J. H.: 'Calculation of Triode Constants', *ibid.*, 1939, **18**, p. 33.
(5) AYER, R. B.: 'The Use of Thoriated Tungsten Filaments in High Power Transmitting Tubes', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 591.
(6) EAGLESFIELD, C. C.: 'Triodes with Square Mesh Grids', *Wireless Engineer*, 1942, **19**, p. 447.
(7) MÜLLER, W.: 'Die Siemens U.K.W. Fernseh Sende-röhren', *FTZ*, 1952, **5**, No. 1, p. 3.
(8) POHL, W. J.: 'Mutual Heating in Transmitting Valve Filament Structures', *Proceedings I.E.E.*, Monograph No. 158 R, December, 1955 (**103** C, p. 224).
(9) ASHDOWN, G.: 'The Vapotron', *Electronic Engineering*, 1953, **25**, p. 378.
(10) British Patent Application No. 24073: 1956.
(11) British Patent Application No. 23986: 1956.
(12) British Patent Application No. 4976: 1957.
(13) KÖPKE, H., and UREDAT, E.: 'Belastungsverhältnisse und Eigenshaften, von Senderöhren im Industriegenerator', *A.E.G. Mitteilungen*, 1955, **45**, p. 523.
(14) APPS, L. T.: 'Class C Amplifier and Oscillator Design,' *Electronic Engineering*, 1955, **27**, p. 323.

(10) APPENDIX

(10.1) Approximate Expression for Grid Flow Angle

The grid flow angle ϕ is given by

$$\cos\frac{\phi}{2} = \frac{V_b}{V_s}$$

With the anode flow angle of 140°, $\phi/2$ is generally less than one radian (i.e. $\phi < 115°$). In the described range of triodes with high slopes per unit area of cathode, ϕ is in the region of 90°. Thus the approximation obtained by expanding

$$\cos\frac{\phi}{2} = 1 - \frac{(\phi/2)^2}{2}$$

is accurate enough for practical purposes. (The error is approximately 1% at 80°, and 4% at 120°.)

This gives $\phi = 2\sqrt{2}\sqrt{\frac{V_g}{V_s}}$ radians.

From the curves in Reference 1, the approximation that I_{1g}/I_g is linear will be seen to be permissible for flow angles up to 120°.

It is proposed to use a value $I_{1g}/I_g = 0\cdot17$, which corresponds to $\alpha = 1\cdot75$ (see Reference 1).

This relation is used in the derivation of a readily usable expression for the drive and output powers, eqns. (5) and (6).

(10.2) Grid-Dissipation Measurement on Practical Circuits

The grid dissipation may be measured on a practical oscillating circuit by measuring I_{dg}, the mean grid current, and V_s, the grid voltage swing, with a valve voltmeter, and using eqn. (7).

Replacing I_g in eqn. (7) by I_{dg}, and assuming a half-sine-wave current flow, one obtains an expression independent of flow angle:

$$\frac{I_{dg}}{I_g} = \frac{2}{\pi}\frac{\phi}{2\pi} = \frac{\phi}{\pi^2}$$

i.e.

$$I_g = I_{dg}\frac{\pi^2}{\phi}$$

$$P_g = \frac{\phi}{4\pi}V_gI_g = \frac{\pi}{4}V_gI_{dg} = 0\cdot79V_gI_{dg}$$

where V_g is obtained from V_s by subtracting the bias voltage, $V_b = I_{dg}R_b$.

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING) JULY 1957

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: "Reflections from the Ionosphere," *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* "Voltampere, Phone, London."
Printed by Unwin Brothers Limited, Woking and London.